



University  
of Glasgow

Mickowska, Anna (2019) *Evaluation and improvement of the structural mimicry of gp120 mimics*. PhD thesis.

<https://theses.gla.ac.uk/41164/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>  
[research-enlighten@glasgow.ac.uk](mailto:research-enlighten@glasgow.ac.uk)

# Evaluation and improvement of the structural mimicry of gp120 mimics

MSc. Anna Mickowska

Thesis submitted in fulfilment of the requirements for the degree of  
Doctor of Philosophy.



University of Glasgow | School of Chemistry

School of Chemistry  
College of Science and Engineering  
University of Glasgow

WestCHEM 

The logo for WestCHEM, featuring the word 'West' in red and 'CHEM' in black, followed by a graphic of four hexagons (one red, three grey) arranged in a cluster.

August 2018

## Abstract

Despite 30 years of research dedicated to the development of an HIV-1 vaccine, a successful candidate capable of elicitation of neutralizing antibodies against the virus has not yet emerged. The Env gp120 mutation rate is very rapid, however in order to maintain its binding-site towards CD4 receptor, certain crucial residues must be conserved within the viral strains, making it an interesting therapeutic target.

This thesis describes the synthesis of discontinuous gp120 epitope mimics with a novel cyclisation linker, which was applied to improve the solubility and purification of the constructs. For the mimicry of gp120, cyclic peptides mimicking the discontinuous epitopes were assembled on a molecular scaffold. The synthesized linear and cyclic peptides, and final constructs with cyclic peptides assembled on a molecular scaffold were analysed by  $^1\text{H}$  NMR and CD spectroscopy. These spectroscopic methods allowed to gain a deeper understanding of the secondary structure of the synthesized compounds. It was found that cyclisation constrained the peptides, when compared to their linear counterparts. The NMR spectra of the final gp120 mimics were superimposed on each other and it was found, that two out of three compounds were almost identical. CD-spectrometry confirmed the results obtained by NMR.

Since binding studies of the synthesised gp120 mimics required access to CD4 and gp120 proteins, therefore approaches towards expression, purification and characterisation of these proteins were made. However, the challenging refolding of CD4D12 resulted in obtaining a heterogenous mixture of properly folded and misfolded protein. The expression of gp120 in mammalian cells was low-yielding and this protein could not be produced in an amount needed for binding studies.

A reliable and reproducible technique was needed to evaluate binding of gp120 mimics to CD4 receptor. SPR was chosen to study the kinetics of the constructs. Careful optimisation of the experimental conditions yielded in a highly reproducible method for the evaluation of gp120 discontinuous mimics.

## Declaration

I declare that, except where explicit reference is made to the contribution of others, the substance of this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

---

Anna Mickowska

---

Prof. R.M.J. Liskamp



# Table of Contents

Abstract .....	ii
Declaration .....	iii
Table of Contents .....	iv
List of Figures .....	vi
Acknowledgements .....	xii
List of abbreviations .....	xiv
1 Introduction .....	1
1.1 A brief history of vaccination .....	1
1.2 Human Immunodeficiency Virus .....	1
1.3 Protein-protein interactions .....	10
1.4 Molecular scaffolds .....	22
1.5 Peptides and their modifications in drug discovery .....	26
1.6 Applications of discontinuous epitopes mimics in drug design .....	29
1.7 Aim of the project .....	31
2 Protein production and characterisation of CD4D12 and gp120 .....	33
2.1 Expression of CD4D12 in <i>Escherichia coli</i> ( <i>E.coli</i> ) .....	33
2.2 Expression of HIV-gp120 in mammalian cells .....	53
2.3 Conclusions .....	59
3 Synthesis of the gp120 discontinuous epitope mimics .....	61
3.1 Introduction .....	61
3.2 Results and discussion .....	64
3.3 CD and NMR-spectroscopy towards structural characterisation of the synthesized gp120 mimics and independent peptide loops .....	90
3.4 Conclusions .....	107
4 Evaluation of the binding of gp120 protein mimics to the CD4D12 by SPR ..	108
4.1 Introduction .....	108
4.2 Results and discussion .....	110
4.3 Conclusions .....	138
5 Conclusions and future work .....	142
6 Materials and methods .....	144
6.1 Molecular biology .....	144
6.2 Protein expression in Bacteria and Purification .....	145
6.3 Protein expression in mammalian cells and purification .....	148
6.4 Protein Characterisation .....	150
6.5 Binding assays - Surface Plasmon Resonance (SPR) .....	153
6.6 Synthesis .....	155

Appendix.....	180
Bibliography .....	221

## List of Figures

Figure 1 HIV-1 entry mechanism. Adapted from [6][10].	3
Figure 2 Architectural and structural envelope-dependent defensive mechanisms of HIV-1. Adapted from [27].	5
Figure 3 X-ray structure of gp120-CD4-antibody complex.	10
Figure 4 Classification of epitopes based on their complexity.	15
Figure 5 Structural classification of protein-protein interactions.	16
Figure 6 Illustration of a principle of surface plasmon resonance. Adapted from [64].	18
Figure 7 An illustration of far UV CD spectra, which are typical for various protein secondary structures.	20
Figure 8 The continuous and discontinuous epitopes.	22
Figure 9 Mimicry of a discontinuous epitope by attachment of the corresponding cyclic peptides onto the molecular scaffold.	23
Figure 10 Examples of (orthogonal) molecular scaffolds used in protein mimicry (PDB entry for miniature protein: 1HYK).	26
Figure 11 CXCR4 chemokine receptor co-crystallised with CVX15 cyclic peptide (PDB entry: 3OE0).	29
Figure 12 The X-ray crystal structure of full length CD4 structure (left; PDB entry: 1WIO) and its first two domains (CD4D12), which are participating in the interaction with HIV gp120 (right; PDB entry: 1GC1).	34
Figure 13 Vector map of pET28a(+) CD4D12 plasmid (left) and the amino acid sequence of CD4D12 (right).	36
Figure 14 DNA electrophoresis of restriction enzyme digest of pET28a(+)-CD4D12 plasmid.	37
Figure 15 SDS-PAGE analysis of test expression levels after induction in <i>E.coli</i> .	39
Figure 16 Analysis of the impact of different stationary phase during affinity chromatography.	41
Figure 17 Protein precipitation observed during dialysis.	42
Figure 18 SDS-PAGE analysis of CD4D12 purification.	42
Figure 19 Analysis of a one-step CD4D12 purification followed by dot blot and UV.	43
Figure 20 SDS-PAGE analysis of one-step CD4D12 purification.	45
Figure 21 a) SDS-PAGE and b) Western Blot analysis of the purified CD4D12.	46
Figure 22 Crystal structure of CD4D12 (PDB entry: 1GC1).	47
Figure 23 Amino acid sequence of the full length CD4 protein (UniProt: #P01730).	48
Figure 24 Far UV CD spectra of expressed CD4D12.	49
Figure 25 Analysis of secondary structure of expressed CD4D12.	50
Figure 26 1D <sup>1</sup> H NMR of CD4D12 protein.	52
Figure 27 a) DNA electrophoresis of restriction enzyme digest of V1Jnstpagp120 plasmid.	55
Figure 28 SDS-PAGE analysis of gp120 expression in HEK293 cells.	56
Figure 29 Western Blot analysis of the gp120 expression post-induction after 24, 48, 72 h.	57
Figure 30 Progress of purification of gp120 by SDS-PAGE and UV.	58
Figure 31 Far UV spectra of gp120 samples: sample 1 (blue), sample 2 (red).	59
Figure 32 X-ray crystal structure of gp120-CD4-antibody complex (PDB entry: 1GC1).	62
Figure 33 Attachment of the cyclic peptides onto the scaffold for the mimicry of the discontinuous epitopes.	63

Figure 34 Structures of cyclisation hinges. ....	70
Figure 35 Synthetic peptide loops for the mimicry of gp120 CD-4 binding site. .	71
Figure 36 Schematic representation of the three final gp120 mimics used in biological and structural studies. ....	73
Figure 37 Schematic representation of the synthesis of construct 22. ....	73
Figure 38 Progress of attachment of the peptide loops on the TAC scaffold to obtain construct 22 as monitored by analytical HPLC. ....	74
Figure 39 Structure of the final construct 22. ....	75
Figure 40 Schematic representation of the synthesis of construct 30. ....	75
Figure 41 Progress of attachment of the peptide loops on the TAC scaffold to obtain construct 30 as monitored by analytical HPLC. ....	76
Figure 42 Structure of the final construct 30. ....	77
Figure 43 Schematic representation of the synthesis of construct 26. ....	77
Figure 44 Progress of attachment of the peptide loops on the TAC scaffold to obtain construct 26 as monitored by analytical HPLC. ....	78
Figure 45 Structure of the final construct 26. ....	79
Figure 46 Schematic representation of the synthesis of construct 26 with different order of loops attachment. ....	80
Figure 47 Comparison of size and hydrophobic character of independent loops with the retention times obtained from analytical HPLC. ....	81
Figure 48 Progress of TIPS removal after 19 h. ....	85
Figure 49 Progress of TIPS removal after 23 h. ....	85
Figure 50 Progress of TIPS removal after 26 h. ....	86
Figure 51 Progress of TIPS removal after 43.5 h. ....	86
Figure 52 Product decomposition after the total of 66.5 h. ....	87
Figure 53 Zooming in on progress of the TIPS deprotection reaction with TBAF·3H <sub>2</sub> O over time monitored by analytical HPLC. ....	88
Figure 54 Progress of the TIPS-removal in one-pot assembly of the discontinuous mimics monitored by analytical HPLC. ....	89
Figure 55 LCMS analysis of the TIPS removal reaction in one pot after 4 h. ....	90
Figure 56 <sup>1</sup> H NMR analysis of linear peptide 1. ....	92
Figure 57 <sup>1</sup> H NMR analysis of linear peptide 2. ....	93
Figure 58 <sup>1</sup> H NMR analysis of linear peptide 3. ....	94
Figure 59 <sup>1</sup> H NMR analysis of Loop 1. ....	95
Figure 60 <sup>1</sup> H-NMR spectra of linear peptide 1 and loop 1 stacked over each other. ....	95
Figure 61 <sup>1</sup> H NMR spectra loop 2. ....	96
Figure 62 <sup>1</sup> H NMR spectra of linear peptide 2 and loop 2 stacked over each other. ....	97
Figure 63 <sup>1</sup> H NMR spectra loop 3. ....	98
Figure 64 <sup>1</sup> H NMR spectra of linear peptide 3 and loop 3 stacked over each other. ....	99
Figure 65 <sup>1</sup> H-NMR stacked spectra of construct 22, loops 1, 2 and 3. ....	100
Figure 66 <sup>1</sup> H-NMR spectra of constructs 22 and 30 stacked over each other. Cyan: construct 22, Red: construct 30. ....	101
Figure 67 <sup>1</sup> H NMR spectra of constructs 22 and 26 stacked over each other. ...	102
Figure 68 Variable temperature <sup>1</sup> H-NMR of the construct 22. ....	103
Figure 69 CD-spectra of the synthesized compounds. ....	106
Figure 70 Superimposition of the CD-spectra of compounds 22 (blue), 26 (green) and 30 (red). ....	107
Figure 71 Illustration of the interacting partners in the SPR experiment. ....	109

Figure 72 Response of the SPR experiment in the form of sensogram. Adapted from <sup>[160]</sup> . .....	109
Figure 73 pH scouting and pre-concentration studies of gp120. ....	112
Figure 74 The process of gp120 covalent immobilisation on the surface of CM5 chip. ....	113
Figure 75 Regeneration scouting. ....	114
Figure 76 Replicate 1 of the interaction of analyte CD4D12 with immobilised ligand gp120 at 25C. ....	115
Figure 77 Replicate 2 of the interaction of analyte CD4D12 with immobilised ligand gp120 at 25C. ....	116
Figure 78 Replicate 3 of the interaction of analyte CD4D12 with immobilised ligand gp120 at 25C. ....	116
Figure 79 Replicate 1 of interaction of analyte CD4D12 <sub>ex</sub> with immobilised ligand gp120 at 25C. ....	119
Figure 80 Baseline and binding level in binding CD4D12 <sub>ex</sub> to immobilised gp120.12 <sub>ex</sub> to immobilised gp120. ....	120
Figure 81 pH scouting and pre-concentration studies of CD4D12. ....	122
Figure 82 The process of CD4D12 covalent immobilisation on the surface of CM5 chip. ....	124
Figure 83 Regeneration scouting of the chip surface with immobilised CD4D12. ....	125
Figure 84 Replicate 1 of the interaction of analyte gp120 with immobilised ligand CD4D12 at 25C. ....	126
Figure 85 Replicate 2 of the interaction of analyte gp120 with immobilised ligand CD4D12 at 25C. ....	127
Figure 86 Replicate 3 of the interaction of analyte gp120 with immobilised ligand CD4D12 at 25C. ....	128
Figure 87 Upper: Schematic representation of loops 1, 2 and 3 structures; lower: schematic representation of the synthesized gp120 discontinuous mimics 22, 26 and 30 structures. ....	129
Figure 88 Unprocessed sensograms of construct 30 binding to: left: the blank sensor surface; right: the immobilised ligand CD4D12. ....	130
Figure 89 Top: Binding to the reference channel (RU) in each cycle, bottom: binding level responses (RU) in each cycle. ....	131
Figure 90 Replicate 1 of performed experiment to measure binding of mimic 22 to the immobilised CD4D12 at 25C. ....	133
Figure 91 Replicate 1 of performed experiment to measure binding of mimic 26 to the immobilised CD4D12 at 25C. ....	134
Figure 92 Replicate 1 of performed experiment to measure binding of mimic 30 to the immobilised CD4D12 at 25C. ....	135
Figure 93 Replicate 1 of performed experiment to measure binding of loop 1 to the immobilised CD4D12 at 25C. ....	136
Figure 94 Replicate 1 of performed experiment to measure binding of loop 2 to the immobilised CD4D12 at 25C. ....	137
Figure 95 Replicate 1 of performed experiment to measure binding of loop 3 to the immobilised CD4D12 at 25C. ....	138
Figure 96 Structures of the cyclisation linkers. Left: N <sub>3</sub> -DBMB; right: (polar hinge) N <sub>3</sub> -TADB. ....	139
Figure 97 Results of protein mass spectrometry of CD4D12. ....	180
Figure 98 HRMS of Tert-butyl 5-(triethylsilyl)-pent-4-ynoate. ....	181
Figure 99 HRMS of Tert-butyl 5-(triisopropylsilyl)-pent-4-ynoate. ....	181
Figure 100 HRMS of 5-(triisopropylsilyl)-4-pentynoic acid. ....	182

Figure 101 HRMS of 5-(triethylsilyl)-4-pentynoic acid.....	182
Figure 102 MALDI-TOF of orthogonally protected trialkyne TAC-scaffold 15. ...	183
Figure 103 HRMS of orthogonally protected TAC scaffold 15. ....	183
Figure 104 LCMS of orthogonally protected TAC scaffold 15.....	184
Figure 105 analytical HPLC of orthogonally protected TAC scaffold 15. ....	184
Figure 106 MALDI-TOF of loop 1. ....	185
Figure 107 LCMS of loop 1. ....	185
Figure 108 analytical HPLC of loop 1. ....	186
Figure 109 MALDI-TOF of loop 2. ....	186
Figure 110 LCMS of loop 2. ....	187
Figure 111 analytical HPLC of loop 2. ....	187
Figure 112 MALDI-TOF of loop 3. ....	188
Figure 113 LCMS of loop 3. ....	188
Figure 114 analytical HPLC of loop 3. ....	189
Figure 115 LCMS of cycloaddition of the first loop in the synthesis of construct 22. ....	189
Figure 116 LCMS of TES removal to obtain product 19.....	190
Figure 117 analytical HPLC of purified compound 19. ....	190
Figure 118 LCMS Cycloaddition of the second loop to obtain product 20. ....	191
Figure 119 analytical HPLC of purified compound 20. ....	191
Figure 120 LCMS of product 21 after TIPS removal. ....	192
Figure 121 analytical HPLC of purified compound 21. ....	192
Figure 122 MALDI-TOF of compound 22. ....	193
Figure 123 LCMS of compound 22. ....	193
Figure 124 analytical HPLC of purified compound 22. ....	194
Figure 125 LCMS of cycloaddition of the first loop in synthesis of compound 30. ....	194
Figure 126 analytical HPLC of crude reaction mixture after cycloaddition of the first loop in the synthesis of compound 30. ....	195
Figure 127 LCMS of TES removal to obtain compound 27.....	195
Figure 128 LCMS of product 28 after cycloaddition of the second loop. ....	196
Figure 129 analytical HPLC of purified product 28. ....	196
Figure 130 LCMS of product 29 after TIPS removal. ....	197
Figure 131 analytical HPLC product 29 after TIPS removal. ....	197
Figure 132 MALDI-TOF of product 30. ....	198
Figure 133 LCMS of product 30. ....	198
Figure 134 analytical HPLC of purified compound 30. ....	199
Figure 135 LCMS after cycloaddition of first loop in the synthesis of compound 26. ....	199
Figure 136 analytical HPLC after cycloaddition of the first loop in the synthesis of compound 26.....	200
Figure 137 LCMS after TES removal to obtain compound 23. ....	200
Figure 138 analytical HPLC after TES removal to obtain compound 23. ....	201
Figure 139 LCMS after cycloaddition of the second loop to obtain product 24. ....	201
Figure 140 analytical HPLC of crude reaction mixture to obtain product 24. ...	202
Figure 141 LCMS of product 25 after TIPS removal. ....	202
Figure 142 analytical HPLC of purified compound 25 after TIPS removal. ....	203
Figure 143 MALDI-TOF of product 26. ....	203
Figure 144 LCMS of product 26. ....	204
Figure 145 analytical HPLC of purified compound 26. ....	204
Figure 146 analytical HPLC: attachment of loop 2 to the TAC scaffold in the synthesis of compound 26 with reverse order of loops attachment. ....	205

Figure 147 LCMS analysis of attachment of loop 2 to the TAC scaffold in the synthesis of compound 26 with reversed order of loops attachment. ....	205
Figure 148 TES deprotection to obtain product 22 in the synthesis of compound 26 with reverse order of loops attachment. ....	206
Figure 149 LCMS analysis of TES deprotection to obtain product 22 in the synthesis of compound 26 with reverse order of loops attachment. ....	206
Figure 150 Progress of attachment of loop 3 (18) to compound 22 to obtain compound 31 in the synthesis of compound 26 with reverse order of loops attachment as analysed by analytical HPLC. ....	207
Figure 151 analytical HPLC of purified compound 31 (in the synthesis of compound 26 with reverse order of loops attachment as analysed by analytical HPLC), which could not be separated from TBTA. ....	207
Figure 152 LCMS analysis of purified compound 31 (in the synthesis of compound 26 with reverse order of loops attachment), which could not be separated from TBTA. ....	208
Figure 153 Progress of TIPS removal from compound 31 to obtain compound 32 in the synthesis of 26 with reverse order of loops attachment as analysed by analytical HPLC. ....	208
Figure 154 analytical HPLC of purified compound 32 (in the synthesis of compound 26 with reverse order of loops attachment as analysed by analytical HPLC). ....	209
Figure 155 LCMS analysis of TIPS removal from compound 31 to obtain compound 32 in the synthesis of compound 26 with reverse order of loops attachment. Crude reaction mixture after 2h. ....	209
Figure 156 analytical HPLC of purified compound 26 in the synthesis with reversed order of loops attachment. ....	210
Figure 157 LCMS analysis of purified compound 26 in the synthesis with reversed order of loops attachment. ....	210
Figure 158 analytical HPLC: attachment of loop 2 to TAC scaffold in an attempt of one-pot synthesis of gp120 mimic. ....	211
Figure 159 analytical HPLC of TES deprotection to obtain compound 23 in an attempt of one-pot synthesis of gp120 mimic. ....	211
Figure 160 analytical HPLC of loop 1 attachment in the synthesis of compound 24 in an attempt of one-pot synthesis of gp120 mimic. ....	212
Figure 161 <sup>1</sup> H-NMR analysis of linear peptide 1 with added TCEP. Water signal was suppressed. ....	212
Figure 162 <sup>1</sup> H-NMR analysis of linear peptide 2 with added TCEP. Water signal was suppressed. ....	213
Figure 163 <sup>1</sup> H-NMR analysis of linear peptide 3 with added TCEP. Water signal was suppressed. ....	213
Figure 164 CD spectrum of linear peptide 1 with TCEP. ....	214
Figure 165 CD spectrum of linear peptide 2 with TCEP. ....	214
Figure 166 CD spectrum of linear peptide 3 with TCEP. ....	215
Figure 167 Replicate 2 of interaction of analyte CD4D12 <sub>ex</sub> with immobilised ligand gp120 at 25C. ....	215
Figure 168 Replicate 2 of performed experiment to measure binding of mimic 22 to the immobilised CD4D12 at 25C. ....	216
Figure 169 Replicate 2 of performed experiment to measure binding of mimic 26 to the immobilised CD4D12 at 25C. ....	217
Figure 170 Replicate 2 of performed experiment to measure binding of mimic 30 to the immobilised CD4D12 at 25C. ....	217

Figure 171 Replicate 2 of performed experiment to measure binding of loop 1 to the immobilised CD4D12 at 25C. ....	218
Figure 172 Replicate 2 of performed experiment to measure binding of loop 2 to the immobilised CD4D12 at 25C. ....	219
Figure 173 Replicate 2 of performed experiment to measure binding of loop 3 to the immobilised CD4D12 at 25C. ....	220



## Acknowledgements

Firstly, I would like to thank my supervisor Prof. Rob M.J. Liskamp for giving me the great opportunity to join his group and work on this project. He always had time for any questions and showed continuous support throughout my time as a PhD student. I will always remember the phrase “Let’s achieve greatness!”.

I would like to thank the Liskamp group for making my time at the university and outside of it a funny and positive environment. Meeting people from so many countries and of different characters was a very interesting and an eye-opening experience. Many thanks to David for finding positive side in everything that surrounded us, for teaching me everything about bikes and sharing many cakes together. I would also like to thank Valery for all his creative and funny ideas and showing me a slightly different perspective to everything. Thank you, Tom, for calling me “your Anna”, for all the funny conversations and singing together in the lab. Also many thanks to Roderick, for being a positive crazy guy, who was always willing to help and talk and occasionally bake cakes together. What an experience that was. Many thanks to Susan for her endless patience and efforts to keep the lab running smoothly. Also, many thanks to Helmus, for being so helpful and full of wisdom. Many thanks to Tobias, for showing me a whole new world of teas and a whole new level of being crazy. Many thanks to Raik, for teaching me a lot of practical things in the lab and “being pretty” together. And all the patience in the world in the efforts to teach me German. I would also like to thank Natalia, for sharing together many evenings with amazing food and wine. Many thanks to Ondra, for being so instructive and helpful.

Many, many thanks to my great friend Frederike. I really wish everyone had in life a friend like her. Always there, always creative, always supportive, always full of helpful ideas. It was my great pleasure to share 4 years in Scotland with Fred, spend our time together on board-games, gossips, food and drinks and last but not least, planning our futures.

Very special thanks to my other great friend, Dominik. For his so 80s soul, for an uncountable number of hikes through Scottish mountains, for sharing passion for nature and so many interesting conversations. And for all the fun we had, for

checking out new ice cream places, and all the cakes and laughter and Balkanaramas. It was a blast.

I would like to thank my best friend, my favourite person, my dear fiancé Krzysiek. There are not enough words to show my gratitude for his continuous support, especially during the most trying time of thesis writing. Thank you for flying to Glasgow for the first two years of my PhD in here, and for moving here for the last two years. Thank you for sharing time together as those were the happiest moments of my life, and I am hoping that there will be many more of them! I never realised until now, that a person can be so understanding and helpful. Thank you for sharing life with me, for staying calm when its needed and all the enchanted moments.

I would also like to thank all my friends back in Poland, especially: Jaś, Kat, Sky, Pizza Dungeon group, Anima and Pompa, who believed in me, visited me in Scotland many times and discovered this beautiful country together with me. My stay in here without your visits and conversations with you would not be as happy as it was. I would also like to send my love to all the Pratchett fans from Poland, with whom I shared so many great moments. May you live in interesting times!

Chciałabym również podziękować mojej Mamie, bez której wyjazd do Szkocji na doktorat nie byłby w ogóle możliwy. Tylko dzięki jej staraniom i nieustajacemu wsparciu moje studia za granicą przeszły ze sfery planów w rzeczywistość. To, gdzie teraz jestem, jest w głównej mierze efektem jej ciężkiej pracy, cierpliwości i miłości. Dziękuję Mamo.

*„Universities are truly storehouses of knowledge: students arrive from school confident that they know everything, and they leave years later certain that they know practically nothing. Where did the knowledge go in the meantime? Into the university, of course, where it is carefully stored and dried”.*

Sir Terry Pratchett

## List of abbreviations

### General

Å	angstrøm
Ac	acetyl
AIDS	acquired immunodeficiency syndrome
anh.	anhydrous
aq.	aqueous
ART	antiretroviral therapy
BOP	(benzotriazole-1-yloxy)tris(dimethylamino)phosphonium hexafluorophosphate
bp	base pair
C	Carboxy terminus of a peptide
calc.	calculated
CCR5	C-C chemokine receptor type 5
CD	circular dichroism
CD4	cluster of differentiation 4
CD4D12	domains 1 and 2 of CD4
CDR	complementarity determining region
CTV	cyclotrimeratrylene

CuAAC	copper catalysed azide-alkyne cycloaddition
CXCR4	C-X-C chemokine receptor type 4
D	domain
DBU	1,8-Diazabicyclo(5.4.0)undec-7-ene
DC	dendritic cells
DCM	dichloromethane
DiPEA	(ethyl)diisopropylamine
DMA	N,N-dimethylacetamide
DMF	dimethyl formamide
DNA	deoxyribonucleic acid
<i>E.coli</i>	<i>Escherichia coli</i>
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
equiv.	equivalent
ESI	electrospray ionisation
Et <sub>2</sub> O	diethyl ether
FDA	US Food and Drug Administration

Fmoc	fluorenylmethyloxycarbonyl
FP	fluorescence polarisation
GdnHCl	guanidine hydrochloride
GSH	reduced L-glutathione
GSSG	oxidised L-glutathione
h	hours
H <sub>2</sub> O	water
HBS-P	HEPES buffer
HCTU	<i>O</i> -(6-chlorobenzotriazol-1-yl)- <i>N,N,N',N'</i> -tetramethyluronium hexafluorophosphate
HFIP	hexafluoroisopropanol
HIV	human immunodeficiency virus
HPLC	high performance liquid chromatography
HRMS	high resolution mass spectrometry
Hz	hertz
IB	inclusion body
IC <sub>50</sub>	half maximal inhibitory concentration
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry

<i>J</i>	coupling constant
K	kelvin
<i>k<sub>a</sub></i>	association rate
<i>K<sub>d</sub></i>	equilibrium dissociation constant
<i>k<sub>d</sub></i>	dissociation rate
kDa	kilodalton
LB	Luria-Bertani medium
LCMS	liquid chromatography mass spectrometry
LDA	litium diisopropylamide
M	molar
m	multiplet
<i>m/z</i>	ratio of mass to charge
MALDI-TOF	matrix-assisted laser desorption ionisation-time of flight mass spectrometry
MDM2	Mouse double minute 2 homolog
MDMX	Double minute 4 protein
mg	milligram
MHC	major histocompatibility complex
MHz	megahertz

min	minutes
mL	millilitre
mM	millimolar
MMP	methyl- $\alpha$ -D-mannopyranoside
MS	mass spectrometry
MSR	microscale thermophoresis
N	amino terminus of a peptide
N <sub>3</sub> -DBMB	azido di(bromomethyl)benzene
N <sub>3</sub> -TADB	azido triazinanetris(2-bromoethanone)
<i>n</i> -BuLi	<i>n</i> -butyllithium
NHS	N-hydroxysuccinimide
nm	nanometre
nM	nanomolar
NMM	<i>N</i> -methylmorpholine
NMR	nuclear magnetic resonance
NRMSD	normalised root mean square deviation
o.n.	overnight
OD <sub>600</sub>	optical density measured at a wavelength of 600nm

oNBS	o-nitrobenzenesulfonyl
PBS	phosphate buffered saline
pH	negative logarithm hydrogen-ion activity ( $-\log_{10}[\text{H}_3\text{O}^+]$ )
PPIs	protein-protein interactions
ppm	parts per million
q	quartet
R	generalised group
R <sub>max</sub>	analyte binding capacity
RNA	ribonucleic acid
rt	room temperature
RU	response unit
s	second
s	singlet
SDS-PAGE	sodium dodecyl sulphate - polyacrylamide gel electrophoresis
SPPS	solid phase peptide synthesis
SPR	surface plasmon resonance
t	time
t	triplet



TAC	triazacyclophane
TB	terrific broth medium
TBAF	tetra- <i>n</i> -butylammonium fluoride
TBTA	tris((1-benzyl-4-triazolyl)methyl)amine
TCR	lymphocyte T receptor
TEAD	transcriptional enhancer factor domain
TES	triethylsilyl protecting group
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TIPS	triisopropylsilyl protecting group
TLC	thin layer chromatography
$t_R$	retention time
UV	ultraviolet
YAP	Yes-associated protein
YT	yeast extract tryptone medium
$\mu\text{g}$	microgram
$\mu\text{M}$	micromolar
$\mu\text{mol}$	micromol

**Amino acids**

Ala	A	L-alanine
Arg	R	L-arginine
Asn	N	L-asparagine
Asp	D	L-aspartic acid
Cys	C	L-cysteine
Gln	Q	L-glutamine
Glu	E	L-glutamic acid
Gly	G	L-glycine
His	H	L-histidine
Ile	I	L-isoleucine
Leu	L	L-leucine
Lys	K	L-lysine
Met	M	L-methionine
Phe	F	L-phenylalanine
Pro	P	L-proline
Ser	S	L-serine
Thr	T	L-threonine

Trp	W	L-tryptophan
Tyr	Y	L-tyrosine
Val	V	L-valine

# 1 Introduction

## 1.1 A brief history of vaccination

The vaccines developed over the first two hundred years since Jenner's lifetime represent one of the brightest chapters in the history of science. Their striking impact on human's longevity and health is best illustrated by the fact that smallpox, once a deadly disease which is estimated to have killed up to 300 million people during the 20<sup>th</sup> century alone, was eradicated in 1979.<sup>[1]</sup> Global efforts towards eradication of polio has reached its final stages, while successful control of measles makes it another potential target for eradication.<sup>[2][3]</sup> Over the course of 20<sup>th</sup> century other vaccines protecting against once fatal diseases, such as diphtheria, pertussis, tetanus and several other, were developed.<sup>[4]</sup> Pasteur's innovative idea of injecting live-attenuated or inactivated pathogens, has been further developed and a wide range of new methods and approaches are used nowadays. Indeed, the vaccination development has moved away from Pasteur's "three Is paradigm" (isolate, inactivate, inject) towards rational design.<sup>[5]</sup> Despite these achievements, infections such as malaria, tuberculosis and HIV, still pose a great threat, and development of vaccines against them have been so far unsuccessful. Current challenges include development of novel, effective vaccines that would target these complex pathogens, be cost-effective and could be used widely enough to achieve their maximum impact.<sup>[2]</sup>

## 1.2 Human Immunodeficiency Virus

A chronic infection with human immunodeficiency virus-1 (HIV-1) is the main cause of the acquired immunodeficiency syndrome (AIDS) - one of the most ravaging pandemics in the human history.<sup>[6]</sup> The HIV epidemic emerged as a result of zoonotic transmission of the simian immunodeficiency virus (SIV) from African primates. SIVs crossed species from chimpanzees and sooty mangabeys to humans and adapted to yield HIV-1 and HIV-2.<sup>[7]</sup> It is estimated that in 2017 approximately 37 million people were living with HIV/AIDS worldwide and that more than 40 million people have died from AIDS-related illnesses since the start of the epidemic in the 1980s. In 2017 21.7 million people had access to the antiretroviral therapy (ART), resulting in decreased mortality rates globally (numbers based on UNAIDS fact sheet 2018).<sup>[8]</sup> Although, due to ART, it is now possible to live with

HIV, the treatment is lifelong, costly and does not eradicate the infection due to a latent viral reservoir. Thus, prevention of HIV onset by a prophylactic vaccination would be extremely beneficial. Despite enormous worldwide efforts directed to the discovery of both preventive and therapeutic vaccines this goal still remains elusive.

### 1.2.1 HIV entry – the fundamentals

As any other virus, HIV can reproduce only by employing host cells machinery. The replication cycle begins with the adhesion of the virus to the host cell, followed by fusion of the cell membrane with the viral envelope (Env) and release of the HIV capsid into the cell. This series of complex protein-protein interactions can be divided into three phases (**Figure 1**).

The entry process begins with binding of the envelope spikes to their primary receptor on the surface of T lymphocytes, CD4. Upon binding, gp120 undergoes conformational rearrangements of V1/V2 and subsequently V3 that allow binding to the co-receptor - either C-C chemokine receptor type 5 (CCR5) or C-X-C chemokine receptor type 4 (CXCR4). Moreover, a bridging sheet, a structure comprised of two double-stranded  $\beta$ -sheets that were previously spatially separated, is formed because of the binding to CD4. Binding of the co-receptor then triggers fusion of the cell and viral membranes. During this process, the fusion peptide (FP) of gp41 inserts into the host cell membrane, followed by 6-helix bundle formation that pulls together the membranes of both cells. This completes the membrane fusion, and results in the formation of a fusion pore, which is followed by the delivery of the viral contents into the host cell cytoplasm.<sup>[6]</sup>

Once inside the cell, the viral enzyme reverse transcriptase (RT) converts the viral RNA into DNA, which then can be fused with the genome of the host (provirus). During cell division, both DNA of the host and proviral DNA are replicated. The proviral DNA serves as a template for transcription of viral RNA which afterwards is translated to yield the viral proteins. A final infectious virion is formed after viral-particle budding and release.<sup>[6][9][10]</sup>

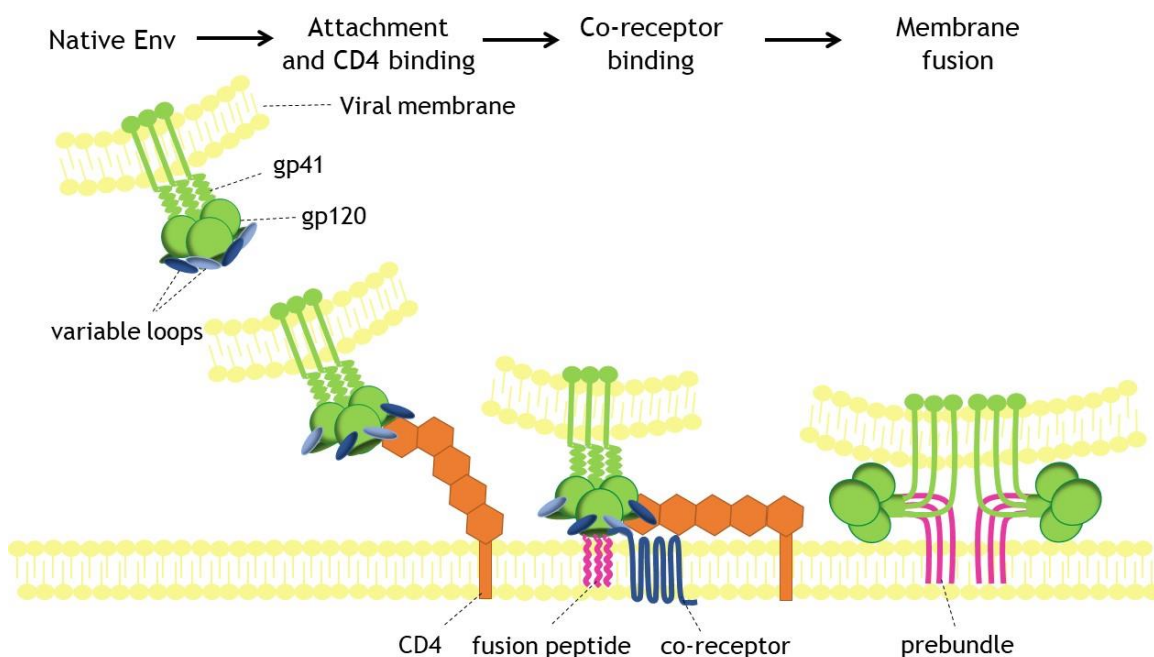


Figure 1 HIV-1 entry mechanism. Adapted from [6][10].

### 1.2.2 Structure of gp120

HIV is an enveloped virus, which contains two single-stranded positive-sense RNA copies inside of its capsid. The glycoprotein gp120 subunit is a significant part of the envelope spikes on the HIV-1 surface.<sup>[11]</sup> The envelope spikes are formed by trimeric heterodimers and consist of three subunits of gp120, a 120 kDa glycoprotein present on the outer surface of the virus, and three subunits of the transmembrane glycoprotein gp41. The complex is formed by cleavage of a polyprotein precursor - gp160.<sup>[12]</sup> These two Env glycoproteins are essential to initiate the infection process of the virus by binding to the CD4 receptor, however, it is gp120 that participates in the binding event directly.

Gp120 is a heavily glycosylated glycoprotein, it is estimated that approximately half of its molecular weight is comprising *N*-linked glycans (20-35 *N*-glycosylation sites) with a small contribution of *O*-linked sugars. Biochemical and mass spectrometry studies have revealed that the majority of these glycans comprise the mannose types - Man-5- and Man-9. The glycans contribute to the correct folding of the protein and shield the immunogenic epitopes of gp120.<sup>[13][14]</sup>

Gp120 contains five relatively conserved domains (C1-C5) which are interspersed by five variable loops (V1-V5), named for their relative genetic heterogeneity.

Regions V1-V4 form surface-exposed loops comprising disulphide bridges at their bases. The variable regions are responsible for co-receptor binding (especially V3) and play role in evasion from the immune system. The variable regions V1/V2 and V3 shield the conserved CD4-binding site on gp120. The role of V4 and V5 is unclear, however V4 region is probably needed for the correct folding of gp120.<sup>[15]</sup>

Highly conserved cysteine residues present in the primary structure of gp120 contribute to the disulphide formation that is crucial to develop the correct tertiary structure.<sup>[12]</sup> Conserved residues, specifically C1, C3 and C4 determinants form the CD4 binding-site. Structural studies performed by Kwong et al.<sup>[16]</sup> have identified the amino acid sequences of the discontinuous loops that form the CD4 binding-site. These are: <sup>454</sup>LTRDGGK<sup>460</sup>, <sup>424</sup>INMWQEVGKA<sup>433</sup>, <sup>365</sup>SGDDPEIVT<sup>373</sup>. The high degree of conservation in the CD4 binding-site makes it an attractive target for broadly cross-reactive neutralization studies. Mimicry of the recognized discontinuous epitopes is an interesting starting point in the design of a potential vaccine against HIV-1.

### 1.2.3 HIV evasion techniques and epitope design approaches

Despite extensive global research efforts dedicated to finding an HIV vaccine there still have not been any therapeutic advances. The development of a protective vaccine is an extremely challenging task due to a variety of genetic and structural evasion techniques that HIV employs to escape the host immune system (**Figure 2**). These evasion mechanisms include:

- 1) the genetic diversity with sequence conservation limited to receptor-binding sites;<sup>[17]</sup>
- 2) the highly error-prone replication and remarkable capacity to accumulate mutations allow for rapid emergence of escape variants under selection pressure of antibodies and drugs;<sup>[17]</sup>
- 3) the receptor-binding site is sterically occluded by variable surface exposed loops;<sup>[17][18]</sup>

- 4) the extensive glycosylation of the envelope restricts access to conserved binding-sites (glycan shield);<sup>[17][19][20]</sup>
- 5) the glycan shield allows the virus to masquerade itself as a host protein and modulate interactions with the host immune system;<sup>[17][19]</sup>
- 6) the CD4 binding site is available only for a very short period of time as a transient intermediate between the epitope being completely shielded and displayed just before binding its target. Therefore, antibodies have only a relatively narrow “window of opportunity”;<sup>[21]</sup>
- 7) the presence of non-functional envelope spikes, such as gp120-gp41 monomers, gp41 stumps or uncleaved gp160 precursors, which may divert the immune response from functional targets;<sup>[22][23]</sup>
- 8) the low density of spikes on the HIV-1 Env limits interspike cross-linking by antibodies;<sup>[24][25]</sup>
- 9) the conformational masking - the antibodies against the CD4-binding site encounter a high energetic barrier not found in any other anti-gp120 antibodies.<sup>[26]</sup>

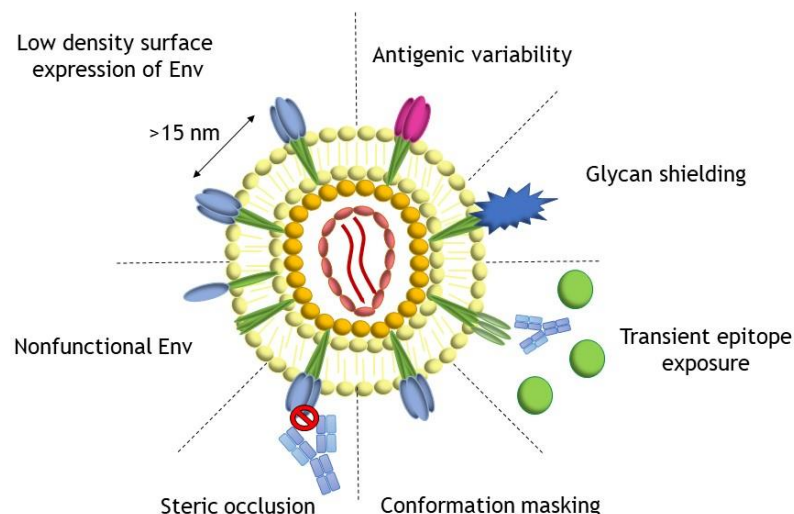


Figure 2 Architectural and structural envelope-dependent defensive mechanisms of HIV-1. Adapted from <sup>[27]</sup>.

These elaborative defensive mechanisms impede the design of an effective vaccine against HIV because of several reasons. The most evident one is that



individuals infected with HIV do not develop the natural immune response, therefore vaccines cannot mimic natural immunological mechanism. During the course of infection, the antibodies elicited are mostly directed against variable and accessible Env loops (immunodominant) rather than less accessible, conserved epitopes, such as the receptor and co-receptor binding sites. Since HIV targets and infects the immune cells it leads to progressive destruction of the immune system. Moreover, the integration of HIV-1 DNA into the host genome results in a latent pool of infected cells which conceal the virus from the immune system. The different strains of the virus present enormous antigenic variability and the degree of antibody affinity maturation required to neutralize HIV-1 is much higher compared with antibodies directed at other viruses.<sup>[28]</sup> Moreover, the vaccine candidates are usually designed based on a unique HIV strain, while an ideal vaccine should not be strain-specific.

Traditional approaches have proved unsuccessful due to the fact, that the immunodominant viral epitopes are mutable while the conserved epitopes necessary for infection are not sufficiently immunogenic.<sup>[29]</sup> Several different strategies were approached to improve the immunogenicity of the epitopes recognized by mAbs. One of the approaches in the design of the epitopes include hyperglycosylation of HIV gp120 to mask undesired epitopes and direct the immune response against conserved epitopes that are important for CD4 binding.<sup>[30]</sup> Other strategies are based on amino acid substitutions<sup>[31]</sup> or stabilization of the CD4-bound state by introducing crosslinks or deleting variable loops.<sup>[32]</sup> An interesting strategy includes immunofocusing, e.g. decreasing the ability of the CD4 binding-site to bind to non-neutralizing antibodies while retaining the capacity to bind to broadly neutralizing mAbs (bnmAbs).<sup>[33]</sup>

Apart from these different epitope-design-based approaches, other tactics towards a therapeutic HIV vaccine were employed (**Table 1**), among them:

- 1) DNA vaccines - designed to encode antigens, which once produced would stimulate an immune response after administration. This approach is safe and has demonstrated promising immunogenicity in animal models. These results encouraged the development of the DNA-based HIV-1 vaccine PENNVAX. After vaccination with PENNVAX with IL-12 plasmid, it was found that most subjects developed CD4<sup>+</sup> or CD8<sup>+</sup> T-cell responses.<sup>[34]</sup>

2) Dendritic cell vaccines - possess an ability to ensure that the antigen reaches the cells most able to activate protective immune responses. However, this approach has the disadvantage of being time-consuming and expensive. An example of dendritic cells vaccine is the DCV2/MANON07-ORVACS which consists of autologous dendritic cells pulsed with whole heat-inactivated HIV-1. It demonstrated immunogenicity associated with a control of viral replication in a phase I clinical trial. After administration of the vaccine the viral load was significantly reduced, which was probably a result of increased HIV-1 specific T-cell responses.<sup>[35]</sup>

3) Subunit vaccines - consist of peptides derived from conserved domains within HIV-1, without introduction of the entire viral particles. As an example of this approach, the Vacc4X subunit vaccine consists of four modified peptides derived from the conserved domains within the HIV-1 p24 group-specific antigen (gag) protein containing major histocompatibility complex (MHC) class I and class II restricted epitopes. It induced T-cell responses in 90% of patients which were associated with reduced viral loads. Seven years after immunization, CD4<sup>+</sup> and CD8<sup>+</sup> T-cell responses were still observed in 95% and 68% of subjects, respectively. These studies provided evidence for long-lasting T-cell memory response to a peptide-based immunotherapeutic candidate for HIV-1 infected patients.<sup>[36]</sup>

4) Viral vector vaccines - especially adenovirus and pox vaccine, created considerable interest due to low cost and ease of manufacture. The principle of action is based on induction of expression of desired genes encoding for antigens, which in consequence would result in prophylactic immunity or T-cell immune responses, that may have a therapeutic effect. The flagship example comes from a 2009 trial of RV144 canarypox-prime glycoprotein 120 boost vaccine, which reported estimated vaccine efficacy of 31.2%. This level of efficacy is insufficient for deployment of the vaccine; however it was the first example of such a response and encouraged research in this field, as it suggested that a preventive vaccine could be made.<sup>[37]</sup>

Altogether, above examples demonstrate that the HIV/AIDS therapeutic and preventative vaccine field is expanding and that new approaches, which differ from traditional methods in the choice of antigen and delivery system are sought. The Holy Grail of HIV vaccine research remains to be the development of

immunogens capable of induction of broadly neutralizing antibodies that could protect against the large number of different strains of HIV circulating globally. Protein- and peptide-based vaccines have the advantage of being safe, cheap and stable. This would be beneficial as the HIV vaccine would have to be distributed among low-and middle-income countries, where they are likely to achieve their maximum impact. Moreover, the relatively simple design allows to characterize methodically and carefully what impact on the immune system they have compared to more complex vaccines that showed several complications that were problematic to identify.

Table 1 Summary of selected anti-HIV therapeutic vaccines in HIV-1 positive and healthy individuals.

Vaccine (type + name)	Vaccine (constituent)	Immune responses	Viral load	Year of publication
<b>DNA Vaccine</b>  <b>PENNVAX</b>	Three plasmids expressing Env, Gag and Pol; plus plasmid containing IL-12	Developed CD4 <sup>+</sup> and CD8 <sup>+</sup> T-cell responses after repeated vaccination	Uninfected volunteers	2013
<b>DC vaccine</b>  <b>DCV2/MANON07-ORVAC</b>	Autologous DC pulsed with whole inactivated HIV-1	Increased HIV-1 specific T-cell responses	Decrease of plasma viral load setpoint $\geq 1$ log was observed in vaccinated groups and was associated with a consistent increase in HIV-1 specific T cell responses.	2013
<b>Subunit vaccine</b>  <b>Vacc4X</b>	Four modified peptides from Gag protein containing MHC class I and II restricted epitopes	Delayed-type hypersensitivity (DTH) reactions. T-cell responses in 80-90% of patients allowing for structured treatment interruption	Significantly improved viral load ratios	2012
<b>Viral vector vaccine</b>  <b>RV144</b>	Recombinant canary pox viral vector engineered to express HIV-1 Gag and Pro and CRF01_AE HIV-1 gp120	Weak evidence of increased CD4 <sup>+</sup> T cell count upon HIV-1 infection	No overall statistically significant reduction in pre-ART viral load upon HIV-1 infection	2009

### 1.2.4 Mimicry of the discontinuous CD4-binding site of HIV-1 gp120

Despite 30 years of research dedicated to the development of an HIV-1 vaccine, a successful candidate capable of elicitation of neutralizing antibodies against the virus has not yet emerged. A number of previously discussed evasive mechanisms of HIV-1 hampers the discovery of an effective vaccine. The Env gp120 mutation rate is very rapid, however in order to maintain its binding-site towards CD4 receptor, certain crucial residues must be conserved within the viral strains, making it an interesting therapeutic target. An ever-growing number of HIV Env X-ray crystal structures in conjunction with neutralizing antibodies and the CD4 receptor, provide detailed structural information, which combined with knowledge about previous attempts in the design of a vaccine candidate, can be used as a starting point. The key binding sites of the conserved gp120 regions have been identified to be three discontinuous peptide sequences. These are:

Loop 1: <sup>454</sup>LTRDGGK<sup>460</sup>,

Loop 2: <sup>424</sup>INMWQEVGKA<sup>433</sup>,

Loop 3: <sup>365</sup>SGDDPEIVT<sup>373</sup>.

The residues contributing in majority to the binding were identified to be Trp<sup>427</sup>, Asp<sup>368</sup> and Glu<sup>370</sup>.<sup>[16]</sup> The conserved binding site is buried within a hydrophobic core of gp120 and is shielded by variable regions (in particular V1, V2 and V3) from efficient antibody recognition. Several studies have shown that during early stages of HIV infection the non-neutralizing antibodies are elicited against the more accessible V1/V2 region making it an immunodominant epitope (**Figure 3**).<sup>[38][39]</sup>

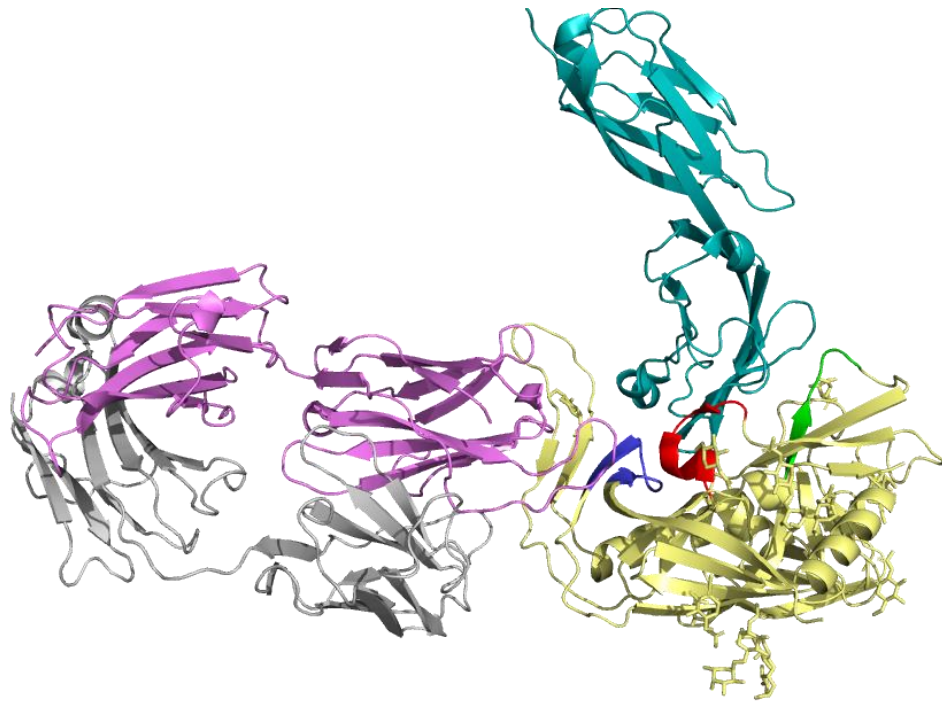


Figure 3 X-ray structure of gp120-CD4-antibody complex. Gp120 (yellow) binds with its discontinuous epitopes (blue, red and green) to CD4 (cyan). On the left-hand side, the antibody (grey and pink) binds to the immunodominant V3 loop. (PDB entry: 1GC1).

Due to its high degree of conservation, the CD4 binding-site is an attractive target for broadly cross-reactive neutralising antibodies. Such antibodies could be elicited by a synthetic mimic (section 1.6) of the gp120 conserved CD4 binding-site in which only the important for the binding event residues would be displayed. Such a mimic could be made by assembly of the cyclic peptides (section 1.5) on a molecular scaffold (section 1.4). In this way the viral evasive mechanisms such as the variable shielding regions and the glycan shield would be omitted. However, reconstruction of the HIV-1 discontinuous epitopes and presenting them in the required conformation at the surface of the scaffold is a challenging task.

The entry of HIV into the cell is dependent on protein-protein interactions (PPIs). PPIs are an important feature of cell biology and an interesting target in drug discovery and will be discussed in more detail in paragraph 1.3.

### 1.3 Protein-protein interactions

Protein-protein interactions (PPIs) can be defined as the physical contacts between two or more proteins that are present in a cell or in a living organism *in vivo*.<sup>[40]</sup> They are involved in a number of key biological processes such as cytoskeletal remodelling, vesicle transport, and signal transduction.<sup>[41]</sup> It is

evident that any dysfunctions in the network of protein interactions might lead to numerous pathological conditions<sup>[42]</sup> e.g. Huntington's disease, osteoporosis, cancer, rheumatoid arthritis, and many other infectious or auto-immune diseases.<sup>[43][44]</sup> Therefore, inhibition or modulation of PPIs provide an attractive opportunity for drug discovery.

### 1.3.1 Protein binding sites as complex structures

Proteins use epitopes (hot spots, binding sites), short peptide sequences that are well defined in three-dimensional conformation, to bind to their targets and elicit biological response. PPIs may occur at the interface of two large protein surfaces that possess electrostatic and structural complementarity to one another.<sup>[45]</sup> The sites involved in these interactions are often flat and devoid of deep pockets that could be targeted by small molecules, which is one of the reasons why they are so difficult to target. The interface area ranges between 1200-2000 Å<sup>2</sup>,<sup>[46]</sup> however it is important to realise, that within this large surface only a subset of residues often clustered in the centre contributes to the binding energy. These hot spots are often shielded by residues that are not directly important to the binding event, however, they play a significant role in occluding bulk solvent from the hot spot, thereby creating suitable dielectric and solvation conditions ('O-ring' theory).<sup>[47]</sup> The resulting non-aqueous environment with decreased dielectric constant favours non-covalent interactions such as hydrogen bonding,  $\pi$ - $\pi$  interactions, hydrophobic interactions and Van der Waals interactions. Analysis of a large protein database of alanine mutants (ASEdb) of heterodimeric protein-protein complexes revealed that tryptophan (Trp), arginine (Arg) and tyrosine (Tyr) are the most common residues found on the interface of protein dimers.<sup>[45][48]</sup> These three amino acids are capable of making multiple types of these favourable interactions (**Table 2**).

Table 2 Trp, Tyr and Arg interactions.

<b>Tryptophan</b>	aromatic $\pi$ -interactions, a hydrogen binding donor, a large hydrophobic surface
<b>Tyrosine</b>	aromatic $\pi$ -interactions, a hydrogen bonding ability, hydrophobic surface
<b>Arginine</b>	hydrogen bond network with up to five H-bonds and a salt-bridge, pseudo-aromatic character, hydrophobic character

Enormous efforts in the pharmaceutical industry and in the academia are being directed at targeting PPIs to inhibit or modulate them. However, this task is very challenging due to several limiting factors: 1) protein binding sites usually have a complex architecture and the crucial interacting residues are often unclear; 2) as mentioned before, the surface areas involved in forming PPIs are large, which is a serious challenge for small molecules to be competitive - the contact surface area they cover is approximately 300-500 Å<sup>2</sup><sup>[49]</sup>; 3) most PPIs occur via discontinuous epitopes thus cannot be mimicked by simple synthetic peptides; 4) the majority of the protein-protein interfaces are relatively featureless, therefore selectivity becomes a difficult task; 5) biological assays that follow enzyme activity are not always applicable, therefore in many cases activity must be detected via more sophisticated techniques that monitor binding directly, e.g. isothermal titration calorimetry (ITC) or surface plasmon resonance (SPR).<sup>[50]</sup> The complex nature of the PPIs inspired researchers to broaden the spectrum of investigated molecules that could potentially modulate the protein interface which was believed to be ‘undruggable’ for a long time. There are three major categories of compounds employed to tackle this problem: fragment-inspired small molecules, protein epitope mimetics, and large recombinant macromolecules.<sup>[42]</sup> Inhibition of PPIs by small molecules has been extensively studied and successful attempts have increased considerably in the recent years with navitoclax and lifitegrast as examples of marketed agents and several other drugs being investigated in clinical trials.<sup>[49][51]</sup> Large recombinant macromolecules, due to their large chemical spaces, can mimic protein interface easier than typical small molecules, and therefore target a wide variety of

different PPIs. However, synthesis and large-scale production of large recombinant molecules is often difficult, time consuming and expensive. On the other hand, peptides and peptidomimetics are believed to have a great potential in modifying PPIs. Rational design based on the identified hot spot sequences allow to focus on the residues important for binding. Moreover, peptides can be chemically modified to obtain the correct spatial conformation and mimic protein structure. An increasing number of peptides and peptidomimetics targeting PPIs is reaching clinical trials.<sup>[52]</sup> A recent example involves a stapled peptide ALRN6924 targeting a tumour suppressor protein p53 via MDMX/MDM2 antagonism. MDM2 and MDMX play a major role in oncogenesis as negative regulators of p53 protein.<sup>[53]</sup> ALRN6924 is currently in Phase I clinical trial in advanced solid tumours and acute myeloid leukaemia.<sup>[54]</sup> However, the hydrophobic character of both peptides and macromolecules, which results in poor solubility and lack of intracellular penetration, presents a serious challenge to their application *in vivo*.

PPIs occur via complex interactions between proteins and can be classified depending on the interaction and structural features of interacting proteins. In order to mimic effectively PPIs in the process of drug discovery, it is important to consider what type of PPIs is targeted.

### **1.3.2 Classification of protein-protein interactions**

Protein-protein interactions can be classified in different ways, depending on the category being analysed.<sup>[55][56]</sup>

#### **I. Depending on the interacting peptide chains<sup>[55][56]</sup>**

PPIs may arise as a result of interaction between identical or non-identical oligomers, for example:

- a) homomeric, e.g. P22 Arc repressor,
- b) heteromeric, e.g. human cathepsin D,



## II. Depending on the protomers stability and lifetime<sup>[55][56]</sup>

Protein that obtain a quaternary structure are built of protomers. Protomer is a subunit of the heterooligomeric protein built of at least two associated polypeptides.<sup>[57]</sup> The protomers vary in their stability and lifetime:

- a) obligate PPIs (strong and long lived) - the protomers are not found as stable structures on their own *in vivo*, e.g. P22 Arc repressor dimer.
- b) non-obligate PPIs (weak and transient) - protomers can exist independently e.g. antibody-receptor, enzyme-inhibitor.

## III. Depending on the complexity of the binding epitopes (Figure 4)<sup>[49]</sup>

PPIs can be classified by the complexity of the epitope interacting with the target protein:

- a) primary structure epitopes - short, continuous, linear peptides, e.g. the linear tripeptide Arg-Gly-Asp (RGD - epitope of fibrinogen) binds to the "I-like" domain of IIbIIIa;<sup>[58]</sup> the linear acetylated histone peptide binds to BRD4 bromodomain.<sup>[59]</sup>
- b) secondary structure epitopes -  $\alpha$ -helix,  $\beta$ -sheet, or extended peptide, e.g. interactions between  $\alpha$ -helix of p53 with MDM2 protein.<sup>[60]</sup>
- c) tertiary structure epitopes - multiple sequences requiring discontinuous binding sites, e.g. binding between interleukin 2 (IL-2) and interleukin 2 receptor  $\alpha$  (IL-2-RA).<sup>[61]</sup>

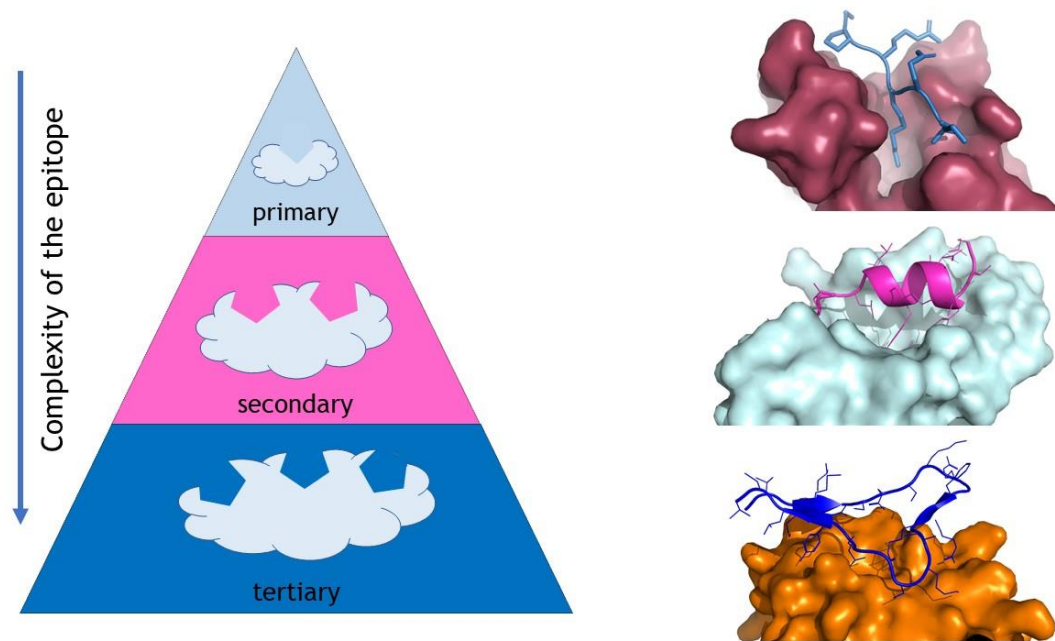


Figure 4 Classification of epitopes based on their complexity. Structures shown are BRD4 (raspberry red) – histone (marine blue) (PDB entry: 2WP1), MDM2 (palecyan) – p53 (magenta) (PDB entry: 1YCR), IL-2 (orange) – IL-2R (blue) (PDB entry: 1Z92). Adapted from <sup>[49]</sup>.

#### IV. Depending on the structural features of interacting proteins (Figure 5)<sup>[56]</sup>

Structurally different proteins can participate in PPIs. Moreover, the PPIs can induce structural change in the proteins. Therefore, PPIs can be classified according to the structural features of interacting proteins:

- a) The interaction via discontinuous epitope between two globular proteins proceeds without substantial structural changes (**Figure 5 a**).
- b) The interaction between two globular proteins involves one or both proteins undergoing a substantial conformational change (**Figure 5 b**).
- c) PPIs involve an interaction between a globular protein and a single peptide chain either with a substantial conformational change, or without a substantial conformational change (**Figure 5 c and d**, respectively).
- d) PPIs involve an interaction between two peptide chains (**Figure 5 e**).

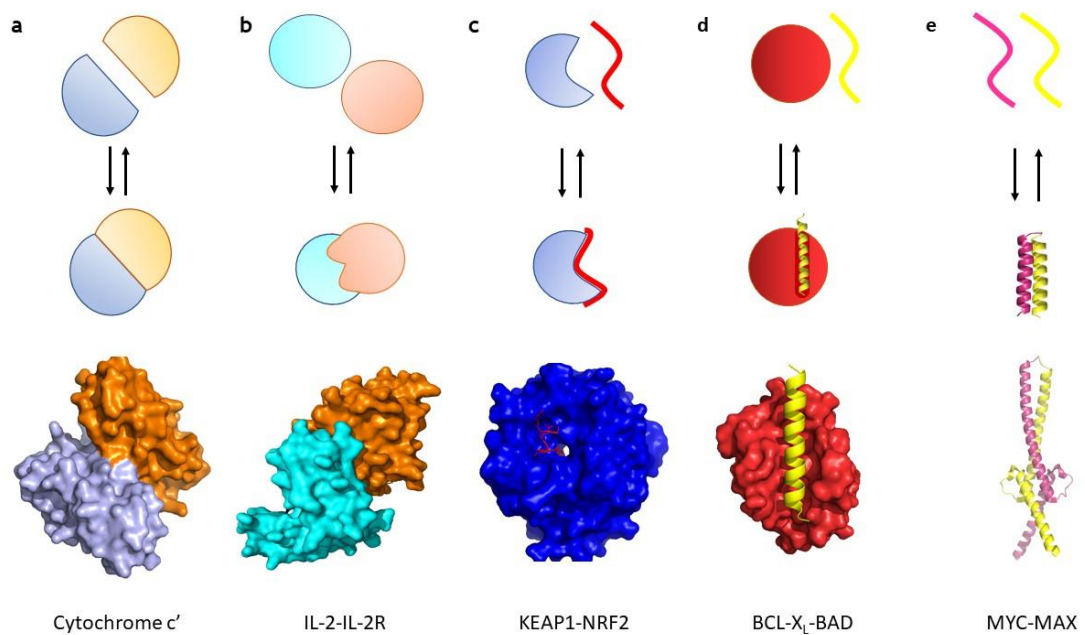


Figure 5 Structural classification of protein-protein interactions.

In the upper part of the figure simplified illustrations are used to represent protein/peptide binding partners, and in the lower part the crystal structures of the exemplary interactions are shown. a) an interaction via discontinuous epitope between two globular proteins without structural changes (PDB entry: 2CCY), b) an interaction between two globular proteins involving conformational change to form binding site (PDB entry: 1Z92), c) an interaction of a rigid protein with a flexible peptide chain (PDB entry: 2DYH), d) an interaction between a flexible globular protein and a peptide (PDB entry: 2XA0), e) an interaction between two peptides (PDB entry: 1NKP). Adapted from <sup>[56]</sup>.

It is evident, that the main function of proteins is formation of complexes with different ligands, such as other proteins or peptides in order to elicit cellular processes that are crucial for a living organism. Formation of these complexes can be studied by various biophysical methods to characterise PPIs, their role and mechanism of action.

### 1.3.3 Biophysical methods for characterizing protein-protein interactions

A variety of experimental methods have been developed to study protein-protein interactions, among them biophysical, biochemical or genetic-based assays. An overview of the key biophysical methods for detection of PPIs is described in this chapter. Each type of a method has its advantages and disadvantages in terms of sensitivity, specificity or sample consumption, which were summarised in **Table 3**.

#### 1) Surface plasmon resonance (SPR)

SPR is a label-free method commonly used in drug discovery. The phenomenon occurs in thin conducting films at the interface between two media (the glass of the sensor chip and the sample solution) of different refractive index. One of the interaction partners (ligand) is immobilised on the surface of the sensor chip (typically a gold chip, **Figure 6**, top) while the other molecule (analyte) is dissolved and injected to flow over the sensor surface. When the chip is illuminated with polarized light at a certain angle of incidence, a proportion of it penetrates through the metal film as an evanescent wave into the medium of lower refraction index. As a result, the energy is absorbed via the evanescent wave field, which is observed as a drop in the intensity of reflected light, which can be detected at a certain angle (SPR angle, **Figure 6**, bottom). The interaction between the ligand on the chip surface and the analyte in the solution results in a change of refractive index at the sensor surface, which is proportional to the change in mass concentration. The alteration in the SPR angle can be measured over time, providing information about the kinetics and the binding affinity of interacting molecules.<sup>[62][63]</sup>

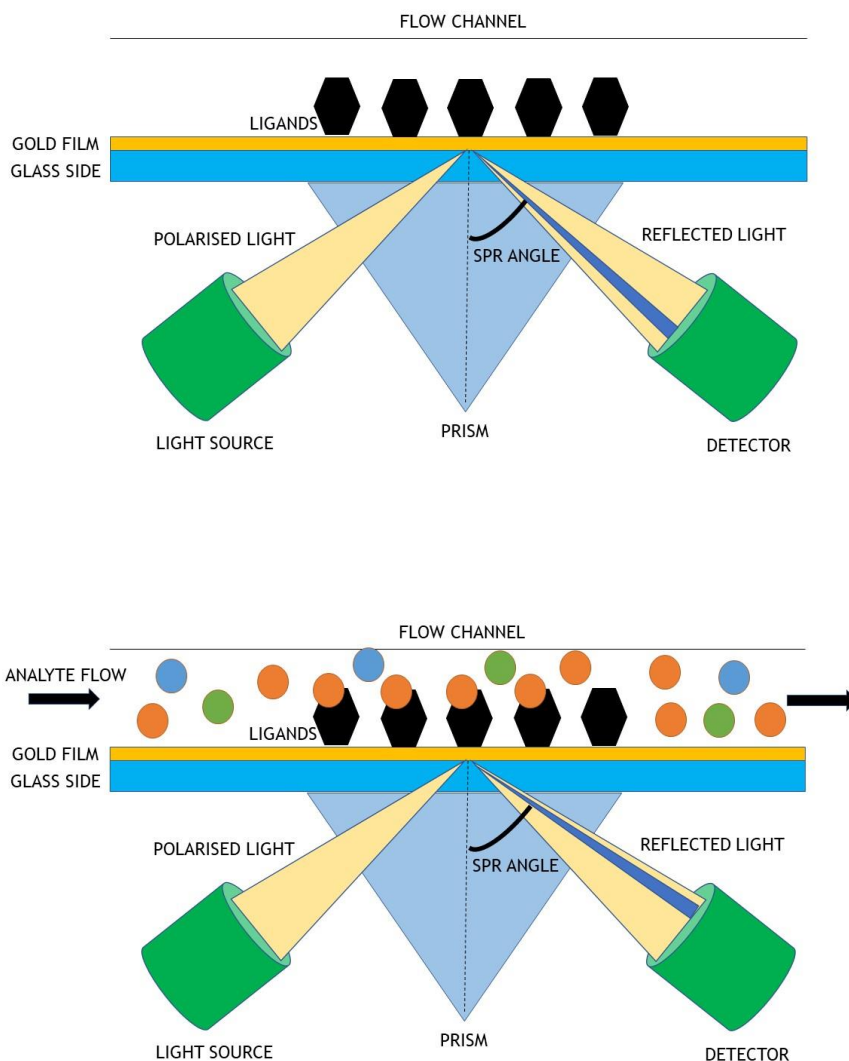


Figure 6 Illustration of a principle of surface plasmon resonance. Adapted from [64].

## 2) Fluorescence polarization (FP)

In FP the movement of fluorescent compounds is observed. Molecules move and rotate in a solution, and their rotation rate is highly dependent on the mass. Therefore, large molecules or complexes of two interacting molecules rotate slower than small molecules or individual binding partners. When a fluorophore is excited by polarized light, it emits light with a certain degree of polarization. Small molecules capable of fast rotation emit depolarized light, while rotation of large molecules/complexes is decreased, therefore the emitted light remains polarized. In an FP assay, typically a small molecule is labelled with a fluorophore and its association and dissociation rates when binding to a large molecule are measured. Complex formation causes an increase in FP signal which can be

measured by a microplate reader. By plotting the FP signal against several concentrations of the labelled molecule a binding curve can be obtained and the dissociation constant ( $K_d$ ) can be determined from the curve.<sup>[64][65]</sup>

### **3) Isothermal titration calorimetry (ITC)**

All chemical reactions proceed with an endothermic or exothermic change in the energy which is detected in ITC. One of the bio-macromolecules (e.g. a protein) is present in a sample cell, the temperature of which is controlled and coupled to a reference cell. When the ligand of interest is titrated in small aliquots, the heat is absorbed or released over the course of titration and can be directly measured. The main parameters measured by ITC include: association constant ( $K_a$ ), enthalpy change ( $\Delta H$ ), reaction stoichiometry ( $n$ ), heat capacity change ( $\Delta C_p$ ), free energy change ( $\Delta G$ ) and entropy change ( $\Delta S$ ).<sup>[66][67]</sup>

### **4) Nuclear magnetic resonance (NMR)**

Nuclear magnetic resonance is a physical phenomenon - a nucleus that possesses spin is placed in a strong magnetic field adopts allowed orientations of different energy. Upon application of electromagnetic radiation, the nuclei absorb the energy and are promoted to a higher energy level. Once the pulse is finished, the nuclei return to its lowest energy level with emission of radiation which can be detected. The frequency at which nucleus resonates depends on its chemical environment. NMR is a powerful technique which allows characterization of the structures of biological macromolecules and to analyse protein-ligand interactions by comparing NMR shifts in their free and bound states.<sup>[68][69]</sup>

### **5) Circular Dichroism (CD)**

Circular dichroism spectroscopy is a useful technique for studying secondary structure of proteins and peptides, as well as protein-protein interactions. The counter-clockwise (left, L) and clockwise (right, R) circularly polarized light is passed through a chiral sample and absorbed to different extents. This difference can be measured and reported in unit of ellipticity ( $\theta$ ) in degrees. CD spectra can be recorded in far UV region (190-250 nm) where absorption arises from the amide groups on the protein backbone thus providing information on the secondary

structure including an  $\alpha$ -helix, a  $\beta$ -sheet and a random coil structures (**Figure 7**). On the other hand, the absorption in the near UV region (260-320 nm) arises from the absorption by aromatic amino acids and disulphide bonds, therefore offering insight into tertiary structure of the sample of interest. The analysis of the differences in the spectra in a free and bound-state provides valuable information about conformation changes accompanying PPIs.<sup>[70][71]</sup>

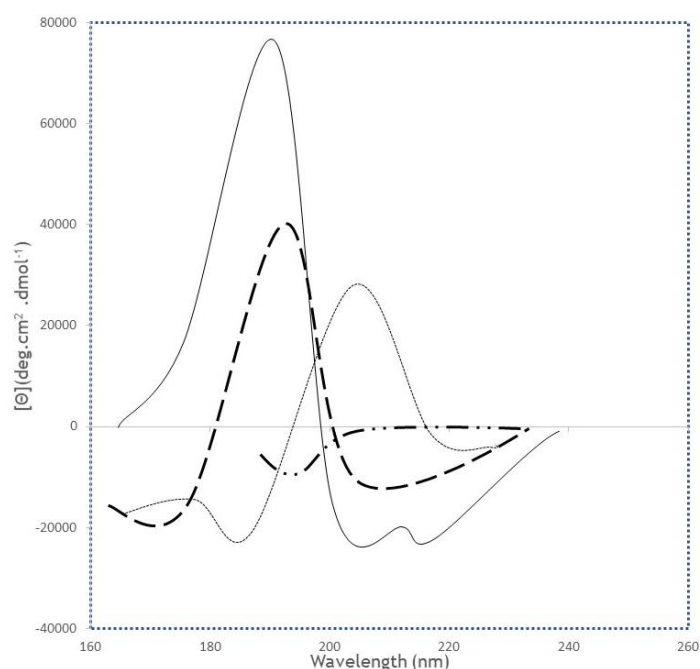


Figure 7 An illustration of far UV CD spectra, which are typical for various protein secondary structures.

Solid curve –  $\alpha$ -helix; long dashes – antiparallel  $\beta$ -sheet; dots – type 1  $\beta$ -turn; dots and dashes – irregular structure. Adapted from <sup>[72]</sup>.

## 6) Microscale Thermophoresis (MST)

MST is a novel and highly sensitive technique that provides information about affinity and thermodynamics ( $K_d$ ,  $n$ ,  $\Delta H$ ,  $\Delta S$ ) of the interaction between binding molecules. In MST the movement of fluorescent molecules (labelled or intrinsically fluorescent) through a microscopic temperature gradient (thermophoresis) is monitored. The movement is highly dependent on various molecular properties, such as size, charge, hydration shell, conformation.<sup>[73]</sup>

Table 3 Advantages and disadvantages of the described biophysical methods for study of protein-protein interactions.

Technique	Advantages	Disadvantages
<b>SPR</b>	label-free, real time kinetic measurement, consumes several $\mu\text{g}$ of sample per chip	immobilisation of the ligand might interfere with the binding event
<b>FP</b>	relatively inexpensive equipment (compared to ITC, SPR), simple protocol (mix-and-read)	autofluorescence and light scattering, anomalous polarisation from aggregation-based non-specific binding might confound the calculations
<b>ITC</b>	a non-destructive technique (does not require immobilisation or labelling of the interacting partners), high-precision and reproducibility (errors typically within 5%)	long preparation time (samples should be dialysed in the same buffer to avoid buffer mismatch), high sample consumption (several hundred $\mu\text{g}$ per binding assay)
<b>NMR</b>	high structural resolution	experimental time, high sample consumption, long time to analyse the obtained spectra, internal protein labelling required, in routine practice the method is limited to proteins with low molecular masses (less than 30 kDa)
<b>CD</b>	label-free, quick assay, non-destructive technique	low structural resolution
<b>MST</b>	fast measurement times, low sample consumption, immobilisation-free	fluorescent labelling required

The described biophysical methods can also be successfully applied to study the activity of the PPI mimics. Such mimics can be made by assembly of the peptides corresponding to the protein epitope on a molecular scaffold, that would display them in a bioactive spatial conformation.



## 1.4 Molecular scaffolds

### 1.4.1 Molecular scaffolds in protein mimicry

As it was discussed in detail in section 1.3, correct three-dimensional presentation of epitopes results in binding to the target protein and elicitation of the biological response. Continuous epitopes consist of a linear amino acid sequence, while more complicated discontinuous epitopes are built of several peptides strands that are relatively far away from each other in the primary structure of a protein, however are close to each other in the three-dimensional conformation of the protein (Figure 8).

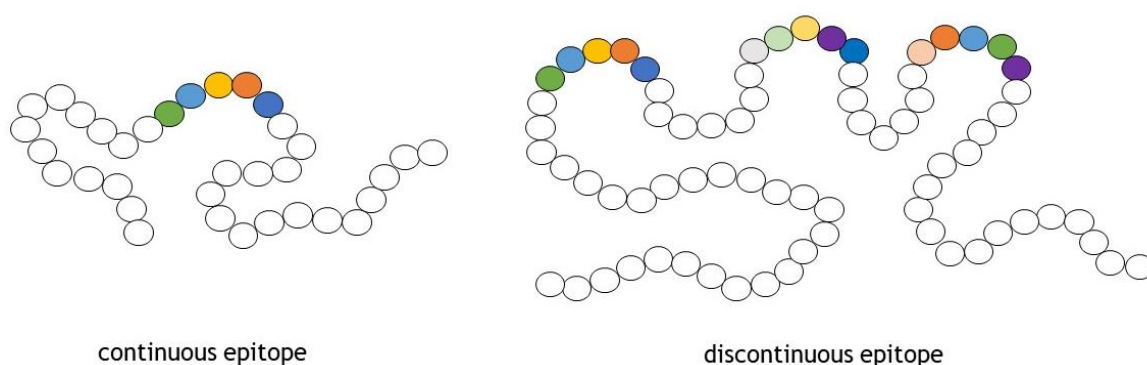


Figure 8 The continuous and discontinuous epitopes.

In theory, the simplest approach for mimicry of discontinuous epitopes is to combine different peptide segments into a single construct. However, short linear peptide motifs are often flexible and disordered, therefore induction or stabilization of the desired secondary structure is required.<sup>[52]</sup> Utilization of a molecular scaffold may represent the best approach to present critical recognition functionalities in the proper bioactive spatial conformation and it is a commonly used tool (Figure 9).

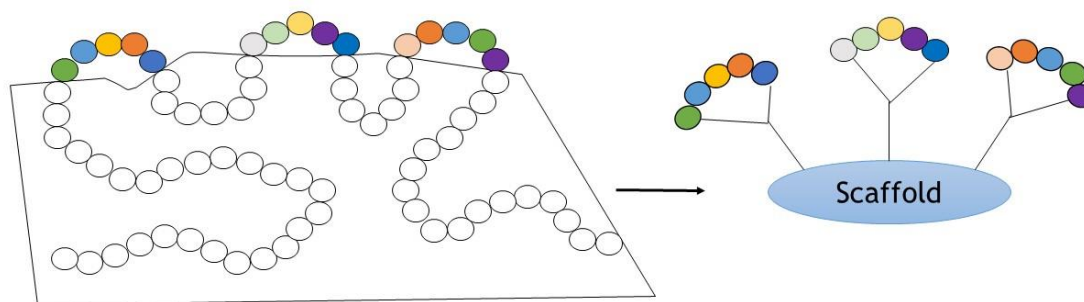


Figure 9 Mimicry of a discontinuous epitope by attachment of the corresponding cyclic peptides onto the molecular scaffold.

A plethora of different molecular scaffolds was reported in the recent years. One of the first examples was the Template-Assembled Synthetic Protein (TASP) scaffold by Mutter et al.<sup>[74]</sup>, which was based on a peptide template (**Figure 10 a**). Initially it allowed for introduction of four identical peptides, however, it was further developed to introduce three different peptide loops by using three different protecting groups (**Figure 10 g**).<sup>[75]</sup> This concept was pushed even further by Dumy et al.<sup>[76]</sup> by implementation of four different protecting groups in four positions to create the Regioselectively Addressable Functionalised Template (RAFT) scaffold, which allowed attachment of a different moiety in each position (**Figure 10 h**).

In another early example a lysine core extended with peptidic arms has been synthesized to create the Multiple Antigenic Peptide (MAP) scaffold (**Figure 10 b**). MAP allows for the introduction of a high density of peptide antigens.<sup>[77][78]</sup> This idea was further explored by Lo-Man et al. who used the lysine core to create a dendrimeric Multiple Antigenic Glycopeptide (MAG) template for the attachment of glycopeptides (**Figure 10 c**),<sup>[79]</sup> which presented high immunogenicity in mice and in non-human primates.<sup>[80]</sup>

Calix[4]arene is an example of a highly pre-organized non-peptidic scaffold, which allows for introduction of four albeit identical peptide loops (**Figure 10 d**). Multiple antigenic units are attached covalently to this multivalent construct, resulting in a better immunogenicity compared to the monovalent construct.<sup>[81]</sup>

Apart from fully synthetic, organic scaffolds, a growing interest has been observed in the utilization of different biomolecules as scaffolds. For example, two

oligoguanosine strands were attached at C- and N-terminus of a peptide via click chemistry and these conjugates underwent the process of self-assembly in the presence of metal ions into a G-quadruplex (**Figure 10 e**). Dimeric assembly of the peptide-oligonucleotide conjugates (POCs) could in principle occur in either one or two faces of the complex depending on the parallel or antiparallel direction of oligonucleotide strands. However, it was shown that parallel topology is selectively formed.<sup>[82]</sup> Another example making use of biomolecules as scaffolds gaining attention is the utilization of miniature proteins - a class of oligopeptides characterized by their short sequence lengths and ability to adopt well-folded, three-dimensional structures. This class of compounds as potential therapeutic agents takes advantage of their small size and can form defined tertiary structures. Functional epitopes can be grafted onto the scaffold to obtain biomolecular recognition of therapeutically relevant targets.<sup>[83]</sup> As an illustration of this approach Silverman et al.<sup>[84]</sup> used agouti-related protein (AgRP), a 4 kDa cysteine-knot protein that has four disulphide bonds and four solvent-exposed loops, as a scaffold grafted with peptides capable of targeting  $\alpha_v\beta_6$  integrins (**Figure 10 f**). Yeast display was used to generate a library of peptides with an RGD-integrin recognition motif that was flanked by randomized residues. In this way, the authors successfully designed highly specific miniature proteins that bound  $\alpha_v\beta_6$  with sub-nanomolar affinity.

Although there are several molecular scaffolds that permit multiple conjugations with peptides, in order to mimic discontinuous epitopes correctly, an orthogonally protected scaffold which allows for the controlled and stepwise introduction of ligands is often needed.

One of the first examples of this new generation scaffolds were the modified TASP and RAFT scaffolds, which are discussed above (**Figure 10 g, h**).

The highly pre-organised orthogonal cyclotrimeratrylene (CTV) scaffold was recently developed in our group (**Figure 10 i**).<sup>[85]</sup> It was designed for subsequent ligation of three different cyclic peptides in order to achieve protein mimicry. It was used lately in the synthesis of highly pre-organised molecules bearing three different peptides mimicking complementarity determining regions (CDR) of the monoclonal antibody Infliximab. The prepared synthetic antibodies exhibited micromolar affinities towards tumour necrosis factor alpha (TNF $\alpha$ ) protein.

Another example of a novel orthogonally protected scaffold developed in our group, which served as a starting point for the synthesis of an HIV vaccine candidate containing three different peptide fragments representing discontinuous epitopes of HIV-1 gp120, is the TriAzaCyclophane (TAC) scaffold. So far, the TAC-scaffold was used successfully in various studies involving papain inhibitors,<sup>[86]</sup> development of a whooping cough vaccine<sup>[86]</sup> and mimicry of gp120 discontinuous epitopes.<sup>[87][88]</sup> Initially it was protected by Alloc, Fmoc and oNBS protecting groups, which allowed for a sequential introduction of peptides through the formation of amide bonds (**Figure 10 j**). However, this scaffold required utilization of protected amino acids, which might result in difficulties with solubility of the construct and further purification. To address these issues, two orthogonally protected and one unprotected alkyne moieties replaced previous protecting groups (**Figure 10 k**). This modification allows the application of Cu(I)-catalyzed azide-alkyne cycloaddition (CuAAC) for the conjugation of unprotected peptide chains modified to contain an azide group. TES and TIPS were chosen as the protecting groups as these can be removed orthogonal to each other, which provides the control over the sequence and position in which peptides are attached.

Recently, a similar approach to the previously mentioned orthogonally protected TAC scaffold was published by Fabre et al (**Figure 10 l**).<sup>[89]</sup> At first, the scaffold also had a free alkyne arm and two arms protected by TES and TIPS. However, it was found that the TES group is unstable in high copper concentrations. Therefore, the TES-protecting group was exchanged by an Fmoc-protecting group, which allowed for coupling to aldehydes, which are often commercially available and are orthogonal with CuAAC chemistry. The final construct can be substituted by two different azides and an aromatic aldehyde, which extends the scope compared to the trialkyne-based scaffold. However, it was found that aliphatic aldehydes cannot be used for coupling, which is a substantial limitation.

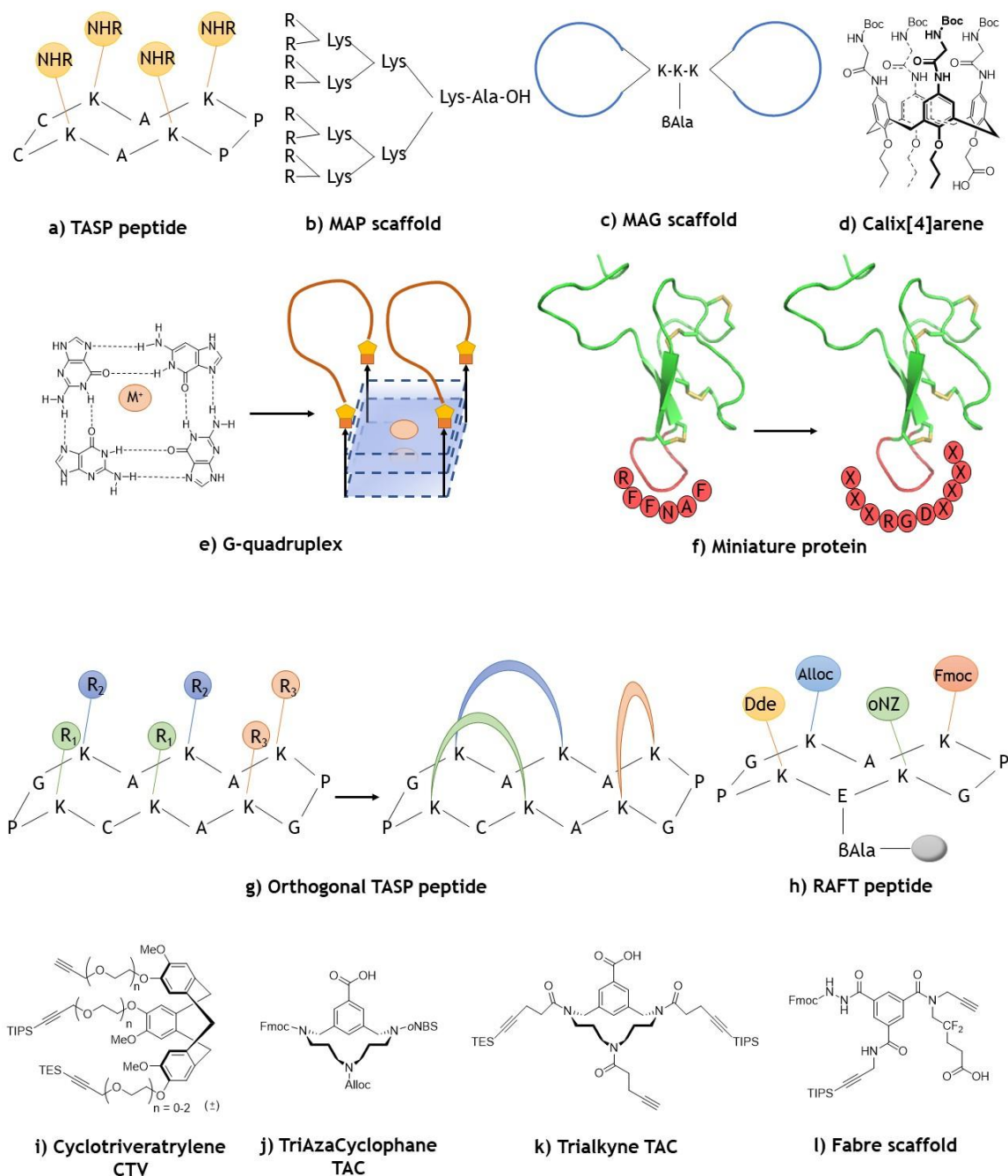


Figure 10 Examples of (orthogonal) molecular scaffolds used in protein mimicry (PDB entry for miniature protein: 1HYK).

## 1.5 Peptides and their modifications in drug discovery

Peptides constitute one of the most promising platforms for drug development in the field of PPIs due to their bio-compatibility, chemical diversity and resemblance to proteins. However, it is well documented that linear peptides have certain features undesirable in therapeutic agents, namely: low stability towards proteolysis, negligible membrane permeability or oral bioavailability,

high clearance and metabolic instability. Moreover, they tend to form a disordered structure in water, whereas their binding target usually recognizes a specific conformation.<sup>[90][91]</sup> To tackle these problems, a variety of chemical modifications has been employed: cyclisation (increased stability), N-methylation (increased membrane permeability and stability), incorporation of unnatural amino acids (increased specificity and stability), attachment of polyethylene glycol polymer (reduced clearance) assorted structural constraints (e.g. disulphide bonds), “stapled” peptides (improved potency and specificity). Chemical modifications of peptides resulted in constructs that were less prone to proteolysis and obtained higher binding affinities and an entropy advantage in receptor binding compared to their more flexible linear counterparts. Reduction of conformational freedom enhances metabolic stability, bioavailability and specificity, thus providing promising lead compounds for drug development. Moreover, the continuous advances in solid-phase peptide synthesis, purification strategies and price reduction of amino acids, have enabled further exploration of peptide-based epitope mimicry.<sup>[91][92]</sup>

Consequently, a growing interest in the peptides was observed in the recent years. Year 2017 noted the highest figure of 46 new drugs approved by the US Food and Drug Administration (FDA) in the last twenty-five years. Among them were five peptides and one peptidomimetic, which together accounted for 13% of the accepted drugs.<sup>[93]</sup> These molecules enable quick access to relatively underexploited higher molecular weight chemical matter that is beyond rule-of-five.<sup>[94][95]</sup> Peptides can address larger and less hydrophobic surfaces involved in many PPIs.

Short synthetic peptide sequences corresponding to the bioactive protein surface do not form thermodynamically stable protein-like structures in water. However, peptides can be induced to fold into protein-like bioactive conformations (strands, helices, turns) by cyclisation. Such constrained cyclic peptides can have protein-like biological activities and potencies, enabling their utilization as biological probes and for development of leads toward therapeutics, diagnostics and vaccines.

### 1.5.1 Cyclic peptides in protein mimicry

Most epitope sequences in a protein have a well-defined secondary structure, such as loops,  $\alpha$ -helices or  $\beta$ -sheets. Several epitopes have effectively been mimicked due to cyclisation of the peptides to obtain a better protein-like structure.

For example, constrained cyclic peptides have been developed to mimic the turn conformation of the RGD sequence. The RGD peptidic motif is the most common motif present in many matrix proteins and is recognized by integrins (cell adhesion proteins). RGD-binding integrins can distinguish between different adhesion proteins because the RGD sequence is presented in slightly different ways.<sup>[96]</sup> On this basis, a series of cyclic pentapeptides with the sequence RGDFV were synthesized and screened leading to the discovery of Cilengitide.<sup>[97]</sup> This drug entered Phase III clinical trials for treatment of glioblastomas, however it did not get final approval.<sup>[98]</sup>

A recent example of protein-protein interactions disruption by a cyclic peptide mimic is the inhibition of the YAP-TEAD complex. YAP (Yes-associated-protein) is a protein responsible for transcription regulation and is known for its oncogenic activity driven by the association with transcription factors from TEAD (transcriptional enhancer factor domain) family.<sup>[99]</sup> In order to inhibit this interaction, firstly the epitope of YAP significant for binding to TEAD was identified to be <sup>81</sup>PQTVPMRLRKLPDSFFKPPE<sup>100</sup>. Based on the crystal structure, residues at positions 87 and 96 were substituted by cysteine and homocysteine and formed a disulphide bond. This resulted in a highly potent cyclic peptide inhibitor.<sup>[100]</sup>

The study by Galsky et al.<sup>[101]</sup> is an example of targeting G protein-coupled receptors (GPCRs) by cyclic peptides. A chemokine receptor, CXCR4 was recently co-crystalized with the CVX15 (**Figure 11**), a 16-residue peptide cyclized by a disulphide bridge, previously identified as an HIV-inhibitor and anti-metastatic agent.<sup>[102]</sup> In order to mimic CVX15 peptide, medium throughput screening and rational design approach were employed and resulted in the identification of the cyclic peptide LY2510924. LY2510924 reached Phase II clinical trials and even though it had an acceptable toxicity profile it did not improve the efficacy of the

first-line standard of care chemotherapy (carboplatin/etoposide) for extensive-disease small cell lung cancer.<sup>[103]</sup>

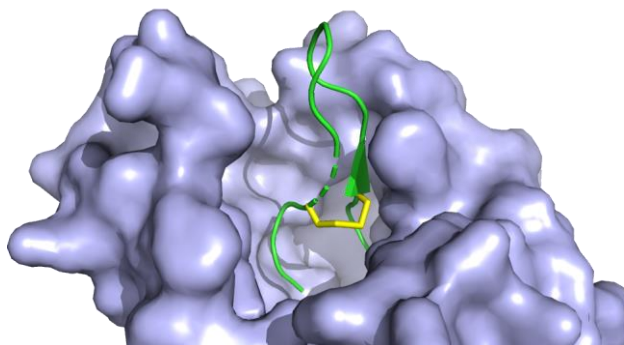


Figure 11 CXCR4 chemokine receptor co-crystallised with CVX15 cyclic peptide (PDB entry: 3OE0).

Cyclic peptides are used more often nowadays in drug discovery. Between 2006 and 2015 nine cyclic peptide drugs were approved by FDA and EMA: four employed in bacterial and fungal infections, three oncology drugs and a drug developed to treat anaemia, this latter one, however, was withdrawn due to safety concerns.<sup>[92]</sup> These data underline the potential that cyclic peptides possess and which should be further explored.

## 1.6 Applications of discontinuous epitopes mimics in drug design

While it would be extremely convenient if linear peptides could be used in the development of vaccines, unfortunately they are rarely targeted by neutralizing antibodies and studies with animals have shown that such antibodies are not elicited.<sup>[104]</sup> The majority of the neutralizing determinants possess quaternary structure, therefore more complex constructs, ones that are capable of obtaining accurate conformation, could be used in the mimicry.<sup>[104]</sup> As an example supporting this hypothesis, the dengue virus Env protein consists of three domains in each protomer. Two protomers form head-to-tail dimers, and a trimer of these dimers forms a raft on the surface of the viral particle. It has been shown, that the human monoclonal antibody (mAb) 1F4 recognizes only a “bent” conformation of the domain I/II region on the dengue Env protein. However, a soluble recombinant Env protein, in which the angle of the domain I/II hinge is more relaxed, does not bind this neutralizing antibody.<sup>[105][106]</sup>



Since protein-protein interactions, and more precisely discontinuous epitopes interactions with proteins, play a crucial role in many diseases, inhibition or controlled modulation of undesired PPIs has become an important target in drug design.<sup>[107][50]</sup>

Peptide-based mimics have gained popularity in recent years as potential therapeutics because of their relatively small size, ease of synthesis, and better bioavailability when compared to proteins. Moreover, as was previously discussed, disruption of PPIs by small molecules is rather difficult. Therefore, utilisation of this family of compounds as discontinuous epitopes mimics is highly sought-after.

### 1.6.1 Protein mimics in vaccinology

Protein mimics have found multiple applications in the field of drug discovery. They can be used as PPIs inhibitors,<sup>[54]</sup> synthetic antibodies,<sup>[108]</sup> anticancer therapeutics,<sup>[109]</sup> antimicrobials,<sup>[110]</sup> anti-inflammatory agents,<sup>[111]</sup> and what is probably one of their most exciting applications - in vaccinology. Live attenuated or killed pathogens, which can be considered as mimics of natural infection, were used to elicit antibody responses *in vivo* and generated many effective vaccines. This approach has been utilised for years, however with certain viruses it was considered unsafe, because of the risk of permanently integrating proviral DNA within host chromosomes, as in the case of HIV.<sup>[112]</sup> Such pathogens could therefore be targeted by careful epitope mimicry in synthetic vaccines.

Vaccines, by definition, are preventive, however in recent years a growing interest was observed in therapeutic vaccines. Prophylactic vaccine is administered to a person who is free of the targeted infection, while in case of therapeutic vaccine it is administered to a person who has already acquired chronic viral infection, against which naturally produced antibodies are ineffective. In such case, the aim is to increase the immune system reactivity to such pathogen.<sup>[113]</sup> Scaffolded protein mimics, especially of discontinuous epitopes, are extremely interesting approach towards synthetic vaccines. If carefully designed to provide correct rigidity and positioning of epitopes and introduce pre-organisation of the peptides so that they obtain the same three-dimensional structures as present in mimicked template epitopes, this approach could lead to new potent vaccines.

## 1.7 Aim of the project

The research described in this thesis aims at the development of a method for the synthesis of discontinuous gp120 epitope mimics and a reliable and reproducible technique to study binding of these constructs to the CD4 protein. Since studies of binding between gp120 mimics and its natural ligand CD4 require access to both proteins, approaches towards expression of these proteins in mammalian and bacterial cells (respectively) are described. Moreover, NMR and Circular Dichroism (CD) spectroscopy studies were performed to obtain a deeper insight into the structural features of the linear peptides, cyclised peptides and entire constructs containing loops mounted on the scaffold.

For the mimicry of discontinuous epitopes, the TAC scaffold was chosen as a platform that should provide the right spatial conformation of the peptides. The TAC scaffold was synthesized based on a literature procedure and one of the steps in the synthesis was optimised. Based on the previously described linear peptide sequences for mimicry of gp120, we describe an optimised synthesis of novel cyclic peptides, containing a recently published cyclisation hinge, which was applied to improve the solubility and purification of the cyclic peptides. Finally, having all the building blocks ready, the cyclic peptides were sequentially introduced on the scaffold, resulting in the complete constructs for gp120 mimicry.

Initially it was decided, that the binding studies of these constructs will be conducted using ITC technique. ITC requires large quantities at high concentrations of the proteins for each assay and using commercial proteins was not financially viable. Therefore, protein expression efforts and their subsequent purification were pursued. Unfortunately, it was not possible to obtain enough proteins in the required purity, therefore the approach towards the binding studies was changed.

Previous approaches towards binding evaluation in the Liskamp group have shown that ELISA assays are highly unreliable and not reproducible.<sup>[114]</sup> Therefore one of the aims of this project was to develop a robust method of evaluation of this class of compounds. In order to achieve this goal SPR was chosen to study the kinetics of the constructs. Careful optimisation of the experimental conditions yielded in a highly reproducible method for the evaluations of gp120 discontinuous mimics.

$^1\text{H}$ -NMR and CD spectroscopy techniques provide valuable information about the folding patterns and three-dimensional conformation of studied molecules.

We describe a first attempt of characterization of the synthesized linear and cyclic peptides as well as entire constructs by these spectroscopic methods to gain a deeper insight into their structure. The results can inspire the design of future constructs to obtain optimised and active compounds.

## 2 Protein production and characterisation of CD4D12 and gp120

### 2.1 Expression of CD4D12 in *Escherichia coli* (*E.coli*)

#### 2.1.1 CD4 – introduction

The main objective of this thesis was to evaluate binding properties of the gp120 discontinuous mimics to the natural ligand CD4 receptor, therefore efforts towards expression of CD4 and gp120 proteins were made.

CD4 (cluster of differentiation 4) is a 55 kDa membrane glycoprotein present on the surface of immune cells such as T helper cells, monocytes, macrophages, Langerhans cells, B lymphocytes, dendritic cells and eosinophils. It plays a major role in the response of the acquired immune system. Under physiological conditions CD4 is a co-receptor of lymphocyte T receptor (TCR) cells and participates in the antigen presentation process. CD4 stabilises the interaction between the antigen presenting major histocompatibility class II (MHCII) cells and lymphocyte T. It also participates in TCR signalling mitigating lymphocyte activation. However, the immune role of CD4 can be corrupted by the HIV-1 virus and it can be used as a port of entry into the cell, which allows further infection by the virus.<sup>[115]</sup> HIV binding and fusion with the host cell is mediated by interaction with the HIV surface gp120 as was described in Chapter 1.

The primary sequence of CD4 consists of 435 amino acids, which constitutes a N-terminal extracellular domain (residues 1-371), a transmembrane segment (residues 373-395) and a C-terminal cytoplasmic tail (residues 396-435).<sup>[116]</sup> The extracellular region is folded into four distinct domains designated D1 to D4. Domains D1, D2 and D4 are additionally stabilised by disulphide bridges, while domains D3 and D4 contain N-linked glycans.

Several studies have shown that the gp120 binding site is located in the first two domains and the minimum essential region participating in this interaction is situated in the first domain (**Figure 12**).<sup>[117]</sup> Within the 98 residues of D1 only 19 have impact on gp120 binding and 13 out of these lie in the region 38-59.<sup>[118]</sup> The

oligosaccharide chains present on domains D3 and D4 have no effect on gp120 binding.<sup>[119]</sup>

One of the aims of this project was to test and evaluate binding properties between gp120 discontinuous epitope mimics and CD4, which was supposed to be pursued by the ITC. ITC requires relatively large quantities (several hundred  $\mu\text{g}$  per binding assay)<sup>[64]</sup> of highly concentrated and purified protein samples thus utilisation of commercial proteins was not financially feasible. Therefore, it was decided that it would be attempted to express and purify the CD4 protein in our laboratory.

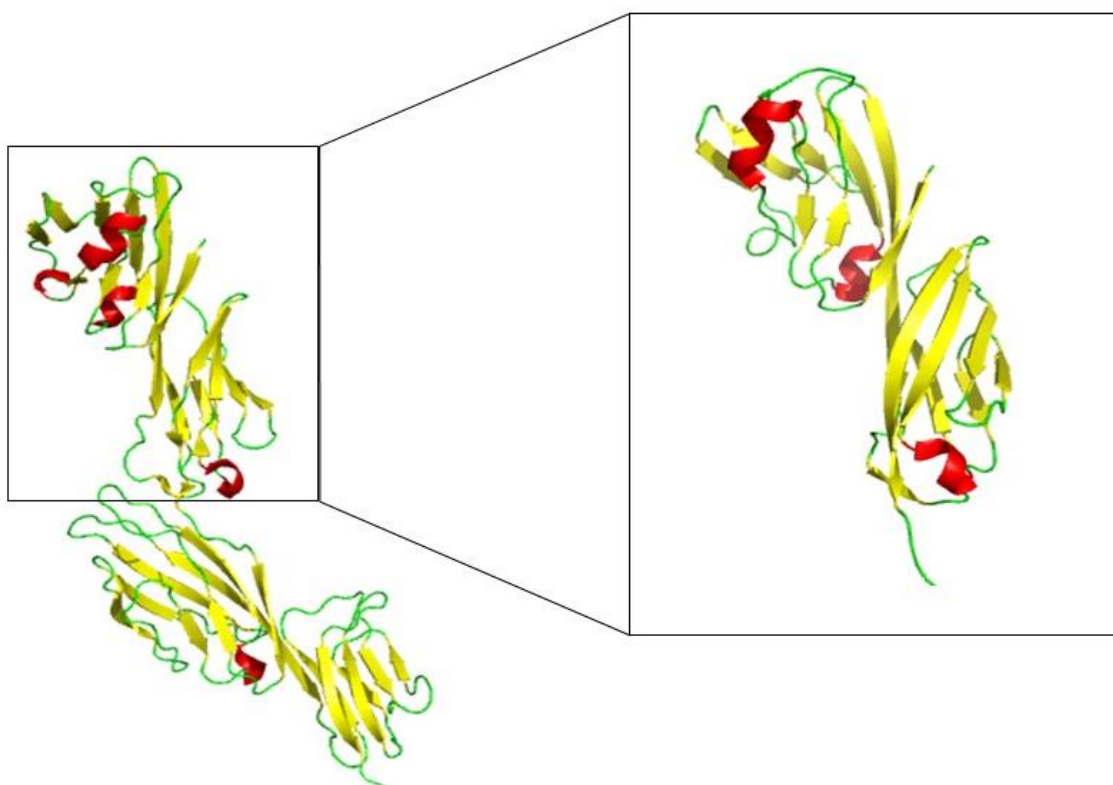


Figure 12 The X-ray crystal structure of full length CD4 structure (left; PDB entry: 1WIO) and its first two domains (CD4D12), which are participating in the interaction with HIV gp120 (right; PDB entry: 1GC1).

### 2.1.2 Overview of CD4D12 expression in *E.coli*

Biologically active recombinant proteins have a variety of downstream applications, such as identification of protein structure and function, characterisation of protein-protein interactions or generation of antibodies. Protein expression in mammalian and insect cells results in biologically active

proteins that contain post-translational modifications, such as glycosylation or phosphorylation.<sup>[120]</sup> On the other hand, production of recombinant proteins in *E.coli* expression system is frequently preferred due to its inexpensive substrates, convenience and high yields.<sup>[121][122]</sup> However, during protein overexpression in *E.coli* inclusion body (IB) formation is frequently encountered. IBs are insoluble products of protein aggregation in a non-native form. While the exact mechanism of IB formation is not yet known, it is commonly believed, that the process of protein expression is faster than its folding into the native form. Hydrophobic regions are exposed on the surface of misfolded or partially folded proteins and interact with other alike proteins, which results in formation of insoluble aggregates.<sup>[123]</sup> For years researchers discarded the insoluble aggregates as an unwanted product obtained during protein overexpression. However, IBs are gaining an increasing attention among researchers due to their mechanical stability that allows harsh physical techniques to lyse the cells, several washing and purification steps, long-term storage stability without loss of protein biological activity and tolerance towards freeze-thawing.<sup>[121]</sup> The major difficulty with proteins being expressed in the form of IBs lies in the solubilisation and subsequent refolding steps to obtain the protein in the native and biologically active form. In order to purify the protein, it is often solubilised first under denaturing conditions (urea, guanidinium chloride - GdnHCl). The high concentration of denaturant is responsible for breakage of non-covalent interactions between molecules and addition of reducing agent helps with prevention of unwanted disulphide bond formation. The most difficult part is the protein refolding step, which often is case-specific and requires trial-and-error to achieve success.<sup>[120]</sup>

Our group contacted professor Varadarajan (Indian Institute of Science, Bangalore) in reference to the plasmid used for expression of the first two domains CD4-D1 and D2 (**Figure 12**, right). DNA plasmid peT28a(+)-CD4D12<sup>[124]</sup> was kindly donated to our group. The peT28(+)-CD4D12 is a plasmid that codes for the first two domains of CD4 protein linked to a polyhistidine fusion tag (His-tag) (**Figure 13**). It is known that CD4D12 is expressed in the form of IBs. Several different protocols for protein overexpression, purification and refolding were tested throughout this work and are described in this chapter.

## 2.1.3 Results and discussion

### 2.1.3.1 CD4D12 expression

#### 2.1.3.1.1 Structure and restriction enzyme digestion of PET28a(+) CD4D12 plasmid

CD4D12 protein expression was carried out using pET expression system as supplied (Figure 13). The gene expression in the pET system proceeds by utilisation of T7 RNA polymerase and bacteriophage T7 promoter.

Modified *E. coli* BL21(DE3) cells which carry the gene coding for T7 RNA polymerase under control of the lacUV5 promoter were used since T7 RNA polymerase is not naturally present in *E. coli*. A pET vector typically consists of: 1) a lac operon (site of *E. coli* RNA polymerase binding) with the gene of interest, 2) an origin of replication (ORI) and the antibiotic resistance gene (KAN(R)) and 3) the lacI gene coding for Lac operon repressor protein. Before induction, the Lac operon repressor protein encoded in the *E. coli* chromosome, inhibits the lacUV5 promoter (within lac operon) by prevention of RNA polymerase binding to the lacUV5 promoter and in consequence halting background expression. During induction, isopropyl B-D-1-thiogalactopyranoside (IPTG) binds to the lac repressor protein in *E. coli* and activates the lacUV5 promoter, which is followed by expression of T7 RNA polymerase. T7 polymerase in presence of IPTG binds to the T7 promoter in the pET plasmid and activates the expression of the gene of interest.

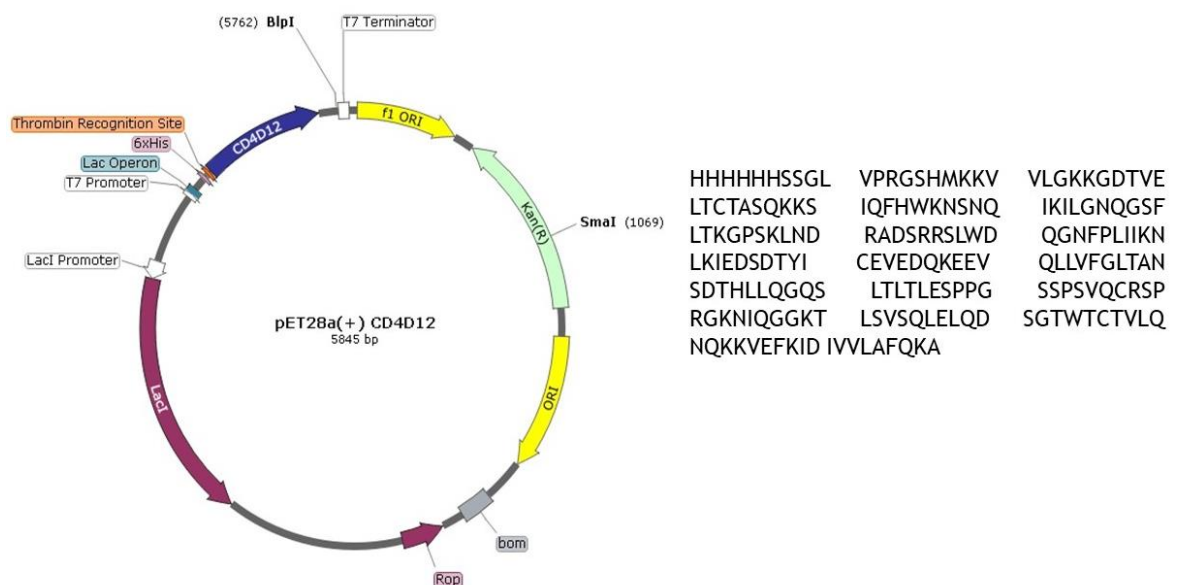


Figure 13 Vector map of pET28a(+) CD4D12 plasmid (left) and the amino acid sequence of CD4D12 (right).

Initially, the donated plasmid was analysed by restriction digestion. The pET28a(+) CD4D12 plasmid consists of 5845 base pairs (bp) (**Figure 13**) and possesses two restriction sites: SmaI and BlnI. A restriction enzyme digestion was performed as described in the Materials and Methods section 6 and samples prior and post the digest were analysed by DNA electrophoresis (**Figure 14**). Prior the restriction enzyme digestion (lane B) the plasmid migrated to a height corresponding to 5,000 bp as expected, however, some bands were visible above, which might be due to formation of a super-coiled plasmid. Restriction enzyme digestion of the plasmid with enzymes SmaI and BlnI produced two DNA fragments of 1152 bp and 4963 bp, as predicted. This result confirmed the size of the plasmid.

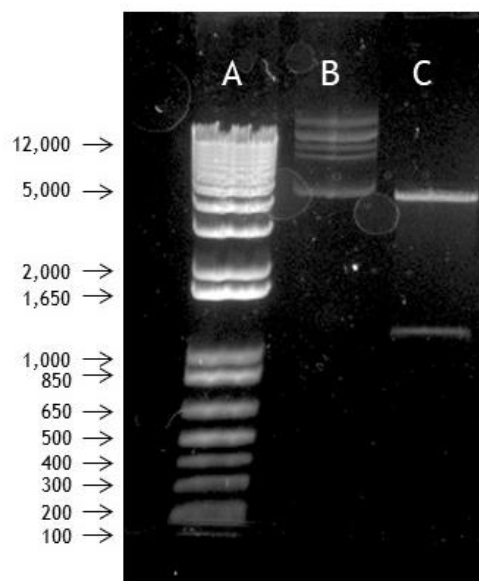


Figure 14 DNA electrophoresis of restriction enzyme digest of pET28a(+)-CD4D12 plasmid. Lane A: 1 Kb Plus DNA ladder, lane B - the plasmid prior to restriction digestion, lane C the cut plasmid after restriction digestion.

#### 2.1.3.1.2 CD4D12 protein expression

The pET28a(+)-CD4D12 vector was transformed into the DH5 $\alpha$  competent cells, which were used to replicate the plasmid. Once a sufficient amount of the plasmid was prepared and purified, the plasmid was further transformed into BL21(DE3) competent cells. The BL21(DE3) strain of *E.coli* is more suitable for protein expression than DH5 $\alpha$  as it is compatible with expression under control of the T7 promoter.

Firstly, different media (2 $\times$ YT, LB, auto-induction LB and TB media) and induction times were tested to optimise the bacterial growth. The comparison was made



based on the size of the pellet of the bacterial culture after centrifugation. The pellets coming from the cultures grown in the auto-induction media and pellets from the cultures induced with IPTG and incubated overnight (compared to shorter incubation times) resulted in similar sized pellets. Therefore, 2×YT medium with manually added IPTG was chosen as the system for protein expression, as it was more economical than utilisation of commercial auto-induction media.

Expression levels after IPTG induction in *E.coli* were monitored by performing test expression. From the inoculated overnight culture, a small sample was collected - this was the uninduced sample. Afterwards the bacterial culture was induced with IPTG until OD<sub>600</sub> reached 7.2 and a further sample was collected - the post-induction sample. The uninduced and induced samples were then lysed and diluted. A fraction of post-induction sample was further used for preparation of following samples: after centrifugation the supernatant (the post-induction soluble fraction) was separated from the pellet, which was resuspended in water, in this way producing the fourth sample - the post-induction insoluble fraction. All these fractions were then analysed by SDS-PAGE. As can be observed on **Figure 15**, prior to the IPTG induction, very weak bands of the proteins were visible (lane B). This might be due to the leaky expression, which is caused by an incomplete repression of the lac promoter leading to low level protein expression prior to the induction.<sup>[125]</sup> After induction many different proteins were expressed (lane C). This post-induction total sample was further divided into soluble (lane D) and insoluble fractions (lane E). The band corresponding to the CD4D12 can be observed in the insoluble post-induction fraction (lane E). The band migrated at between 25 and 35 kDa, while the molecular weight of the two-domain CD4D12 is 21 kDa. The anomalous migration can be caused by the relatively high content (10%) of aspartic acid and glutamic acid residues. Examples in the literature show, that when above 20% of the amino acids in the sequence are acidic, they might repulse the negatively charged SDS, thus migrate slower through the gel, which can lead to the molecular weight differences of even up to 50 kDa between where the band is predicted to migrate and the actual migration position.<sup>[126][127]</sup> In case of CD4D12 the content of acidic residues was not as high and the band migrated only slightly higher than expected. The soluble post-induction fraction did not contain any CD4D12, which was in agreement with the hypothesis, that CD4D12 was overexpressed in the IBs. Moreover, it can be seen that the post-induction

insoluble fraction (lane E) was relatively pure compared to the post-induction soluble fraction.

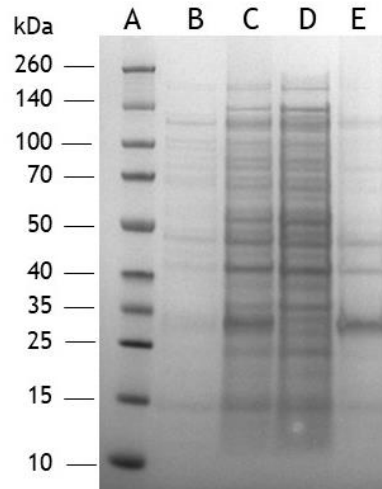


Figure 15 SDS-PAGE analysis of test expression levels after induction in *E.coli*. Lane A: molecular weight marker with values in kDa, lane B: pre-induction total sample, lane C: post-induction total sample, lane D: post-induction soluble fraction, lane E: post-induction insoluble fraction.

### 2.1.3.2 Purification of CD4D12

Purification of the CD4D12 protein consists of two steps: cell lysis, harvest of the inclusion bodies and solubilisation of inclusion bodies under denaturing conditions, followed by  $\text{Ni}^{2+}$ -affinity chromatography.

#### 2.1.3.2.1 Cell Lysis and Solubilisation of Inclusion Bodies

Sonication was employed as a method of cell lysis of all protein samples as described in the literature. <sup>[124][128]</sup> It resulted in cell opening and release of the IBs into the cell lysis buffer. Supernatant was removed by decantation after high speed centrifugation. Afterwards the pellet of IBs was washed several times with a solution of methionine and EDTA. The pellet was washed further with water to remove the EDTA which would interfere with  $\text{Ni}^{2+}$ -affinity chromatography. The insoluble protein aggregates were solubilised in a solubilisation buffer containing high concentration of a denaturant (GdnHCl) and reducing agent ( $\beta$ -mercaptoethanol or TCEP). This process allowed the protein to unfold into its primary structure, preventing formation of undesired disulphide bridges, and to expose the His-tag, which was otherwise buried in the tertiary structure of the protein, which would prevent successful purification. Once the protein was

solubilised, a filtration step was required to remove any insoluble cell debris, thus creating a lysate that could be applied on the  $\text{Ni}^{2+}$ -column.

#### 2.1.3.2.2 $\text{Ni}^{2+}$ -Affinity Chromatography

$\text{Ni}^{2+}$ -affinity chromatography is one of the established methods of purification of recombinant proteins containing His-tag.<sup>[129]</sup> It is widely used and highly reliable. The method is based on the affinity between the imidazole side chain of the histidine tag and nickel ions which are present in the stationary phase. The tagged protein is retained on the column while the untagged proteins are eluted. Once the impurities are removed, the protein can be eluted with an increasing concentration of imidazole (from 20-300 mM), which forces the histidine-tagged protein out of coordination to  $\text{Ni}^{2+}$ . The proteins that have natural affinity towards  $\text{Ni}^{2+}$  are washed away with 20 mM imidazole prior to elution of the desired protein. The buffers (solubilisation, wash and elution) contain NaCl which prevents unspecific binding, and a reducing agent which prevents formation of intra- and/or intermolecular disulphide bridges.

Two procedures were developed for purification of CD4D12. The first one, which was used initially and was not optimised, involved a two-step purification by  $\text{Ni}^{2+}$ -affinity chromatography followed by a one-step dialysis and size exclusion chromatography. Due to limited access to size exclusion chromatography equipment only a small batch of the protein was purified in this way. The protein purified in this way was used for preliminary studies by ITC, which consumed the entire protein stock. The second procedure involved only a one-step  $\text{Ni}^{2+}$ -affinity chromatography and long gradual dialysis in order to obtain correctly refolded protein. This procedure aimed at obtaining the protein in high enough purity after just one step of the purification and focused on the refolding method. The protein produced in this way was used in all characterisation techniques (SDS-PAGE, Western Blotting, NMR and CD-spectroscopy) as well as in the SPR binding studies, unless stated otherwise.

Initially, the test purifications by  $\text{Ni}^{2+}$  and  $\text{Co}^{2+}$ -affinity columns were performed to compare which method was better. It was found that the purity of the eluted CD4D12 was similar, however  $\text{Ni}^{2+}$  ions bound the His-tagged protein slightly better than  $\text{Co}^{2+}$  ions, which resulted in higher amounts of the protein found in the eluted

fractions than in the flow-through. The intensity of the CD4D12 peak was approximately 5× lower during the purification using the  $\text{Co}^{2+}$  resin (200 mAu) than when the  $\text{Ni}^{2+}$  resin was used (1000 mAu), which is shown in **Figure 16**. Therefore, all the purifications were performed using  $\text{Ni}^{2+}$ -packed columns thereafter.

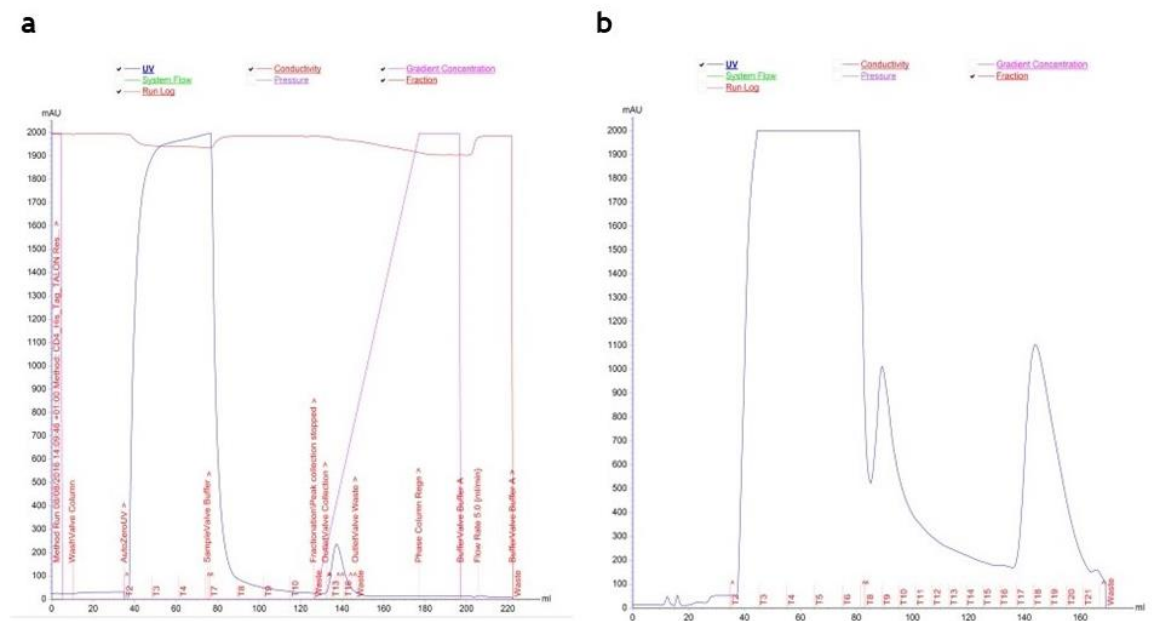


Figure 16 Analysis of the impact of different stationary phase during affinity chromatography. a)  $\text{Co}^{2+}$ -packed affinity column. The broad peak on the left was the flow-through, the peak on the right - CD4D12. b)  $\text{Ni}^{2+}$ -packed affinity column. The broad peak on the left – flow-through, the middle peak – washes, the peak on the right – CD4D12.

The purification of non-denatured protein was also attempted, however it was unsuccessful. It underlined the necessity of exposing the His-tag, which was otherwise buried in the protein and did not bind the resin.

## 1. Two-step protein purification for ITC

Solubilised and filtered protein sample was loaded onto the  $\text{Ni}^{2+}$ -column and washed with 50 mM imidazole wash buffer in order to remove any impurities that might have a natural affinity for  $\text{Ni}^{2+}$ . The protein was then eluted with 500 mM imidazole buffer supplemented with a high concentration of denaturant. After purification, a one-step dialysis to remove GdnHCl was performed. Because of a rapid change of concentration of denaturing agent in the protein sample, formation of a considerable precipitate was observed (**Figure 17**).



Figure 17 Protein precipitation observed during dialysis.

The precipitate was removed by filtration and the soluble fraction of the protein was concentrated and analysed by SDS-PAGE. It was observed, that CD4D12 was already of good purity. Since CD4D12 elution was observed already at the washing step, it was concluded that in future purifications, the imidazole concentration in the wash buffer should be decreased (**Figure 18 a**). Finally, CD4D12 sample was purified by size exclusion chromatography which yielded highly pure protein (**Figure 18 b**), however the amount of CD4D12 produced in this way was very low due to the loss of large part of the sample during the dialysis step.

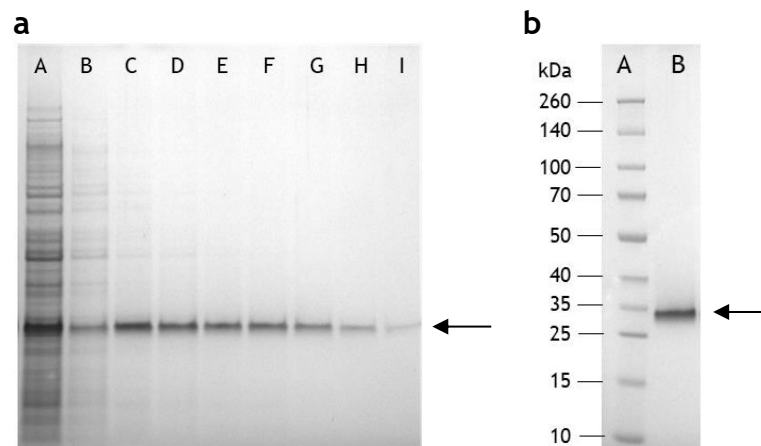


Figure 18 SDS-PAGE analysis of CD4D12 purification.

Left: SDS-PAGE after the first step purification and dialysis of the protein to remove GdnHCl. Lane A: flow-through, lanes B-E: washes, lanes F-I: elutions of the protein. Right: SDS-PAGE of CD4D12 after size exclusion chromatography. Lane A: molecular weight marker with values in kDa, B: purified CD4D12 under reducing conditions.

This method could be optimised by gradual dialysis of the protein sample after the affinity chromatography step, which would circumvent the loss of the protein due to precipitation. Size exclusion chromatography was much preferred as a final purification step as it yielded protein of high purity and it should be pursued in

the future. Refolding of the protein by dialysis was studied in more detail and is described below.

## 2. One-step purification of CD4D12

Solubilised and filtered CD4D12 sample was loaded onto a  $\text{Ni}^{2+}$ -affinity chromatography column and washed with 20 mM imidazole wash buffer and eluted with 300 mM imidazole, as it was observed that CD4D12 eluted at approximately 150 mM imidazole. Purification of the protein was monitored by UV as well as dot blot analysis with an anti-CD4 antibody. This second method was used as a quick test to confirm presence or absence of the CD4D12 in the samples, as SDS-PAGE analysis could not be carried out prior to removal of GdnHCl by dialysis. Dot blot analysis of all the eluted fractions (flow-through, washes, elution fractions) revealed that CD4D12 eluted in all the fractions (**Figure 19**). This might be due to a high concentration of CD4D12 in the lysate and saturation of the  $\text{Ni}^{2+}$ -column. The purest fractions according to the chromatogram were chosen for further steps. This method of purification together with the refolding method described below was used for preparation of protein samples for NMR and CD-spectrometry and SPR binding studies unless stated otherwise.

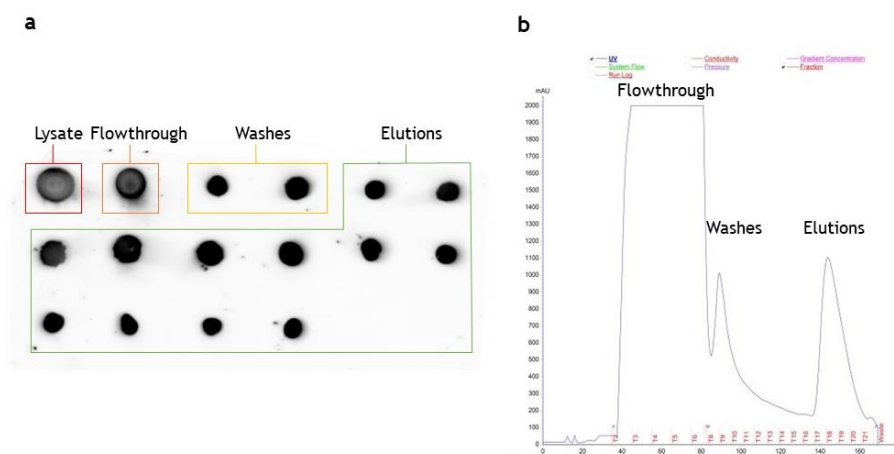


Figure 19 Analysis of a one-step CD4D12 purification followed by dot blot and UV.  
a) dot blot analysis of all the fractions prior (lysate) and obtained after purification (flow-through, washes, elutions); b) the chromatogram obtained after the affinity purification.

### 2.1.3.3 Refolding

Several approaches were utilised in protein refolding to retrieve its native structure and biological activity. One of the traditional methods of refolding which was used in this work is dialysis.<sup>[120]</sup> In a typical procedure, the process of refolding

from the unfolded, primary structure of the protein into its folded, active form is accomplished by removal of denaturing agent. Removal of the denaturant can be performed in one step from high to low concentration of denaturant or in several steps by decreasing the concentration via intermediate concentrations of denaturant. During one step dialysis, the longer the dialysis, the more the concentration of denaturant in the protein sample is decreased and the refolding rate of the protein is increased. However, as the hydrophobic surfaces of the protein are rapidly exposed interactions with each other may cause aggregation of the protein. The rate of misfolding of the protein increases with the number of protein aggregates contacting each other and other protein molecules in the process of refolding. In stepwise dialysis, this problem can be circumvented by gradual removal of the denaturant, so the folding process is slower, and the number of misfolded proteins is decreased. In this way properly-folded protein can be obtained. During the course of this project both techniques were evaluated.

#### **2.1.3.3.1 Refolding of CD4D12**

Following the one-step purification, the refolding of the protein via gradual dialysis was performed. This process was performed according to a slightly modified published procedure.<sup>[130]</sup> After purification, CD4D12 was present in the highest concentration of the denaturant - 6 M GdnHCl. Pure fractions were then dialysed overnight against 3 M GdnHCl buffer, which also contained 10% sucrose as a stabilising agent in the process of refolding, and a redox system consisting of 1 mM reduced L-Glutathione (GSH)/0.1 mM oxidised L-Glutathione (GSSG). The presence of the redox system in the refolding buffer facilitates process of disulphide shuffling, which allows the disulphide bonds to be reduced and oxidised repeatedly until correct folding of the protein is achieved.<sup>[123]</sup>

The next step was to remove the denaturing agent entirely from the buffer and place the protein in the refolding buffer with agents facilitating proper refolding: 10% sucrose, 0.1 mM GSH/0.01 mM GSSG and Na<sub>2</sub>CO<sub>3</sub> to keep the pH at 9.6. High pH is also beneficial in the formation of disulphide bridges, while addition of sucrose stabilizes protein by enhancement of interactions between the additive and the hydrophobic side chains of the denatured protein.<sup>[120]</sup> When changing the refolding buffer from the intermediate concentration of denaturing agent to the

buffer devoid of the denaturant some precipitation was observed. When compared to the one step dialysis, the precipitation in this case occurred to a lesser extent. However, the pH used during refolding was close to the calculated theoretical pI of the protein (9.18, as calculated by the online program [https://web.expasy.org/compute\\_pi/](https://web.expasy.org/compute_pi/)), which meant that the protein net charge was close to 0, and the protein was prone to aggregation. This aggregation was not observed during previous steps, because the buffer was supplemented with denaturing agent, which prevented aggregation and precipitation. However, when the denaturant was removed, the protein started to aggregate and precipitated out of the solution. Future wise, it would be beneficial to lower the pH of the refolding buffers by at least 1 pH unit, so the pH would be still high enough for the bridge disulphide formation, but the protein would be still charged and soluble in the solution.

The last step was an extensive dialysis against PBS buffer at pH 7.4, which was the final buffer in which the protein was stored. No precipitation was observed at this point. The purity of the protein purified in this way was analysed by SDS-PAGE (**Figure 20**). Using this method, (2-3 mg of purified CD4D12 was obtained per 1 L of bacterial cell culture).

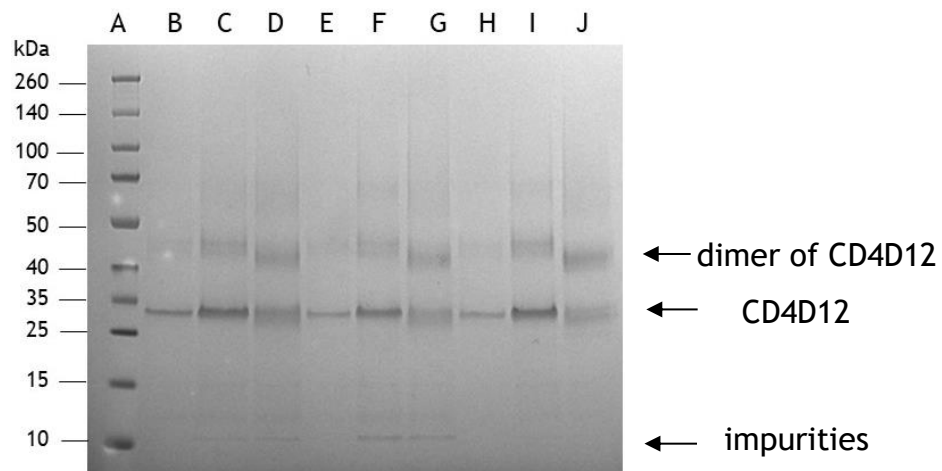


Figure 20 SDS-PAGE analysis of one-step CD4D12 purification.

Lane A: molecular weight marker with values in kDa, Lanes B, C – fraction 1. of the CD4D12 under reducing conditions at concentrations 1 µg·ml<sup>-1</sup> and 2.5 µg·ml<sup>-1</sup>, respectively. Lane D: fraction 1. of the CD4D12 under non-reducing conditions at 2.5 µg·ml<sup>-1</sup>. Lanes E-G and H-J represent fractions 2. and 3., respectively, presented as fraction 1.



### 2.1.4 Confirmation of identity of CD4D12

Upon production of CD4D12 it was needed to confirm that the protein that had been over-expressed was definitely CD4D12, and that it had folded correctly. This was achieved by several methods: Western Blotting, mass analysis after in-gel trypsin digestion, 1D  $^1\text{H}$ -NMR and CD spectrometry. Western Blotting is a commonly used technique to detect specific proteins. SDS-PAGE is used to separate proteins in the sample by mass and then they are transferred from the gel to a membrane. The membrane is firstly incubated with a solution of a primary specific antibody that targets the desired protein, which is followed by incubation with a secondary antibody that targets the primary antibody and is conjugated to an enzyme horseradish peroxidase (HRP). When multiple secondary antibodies bind to one primary antibody, the signal is amplified which allows visualisation.

SDS-PAGE and Western Blotting carried out with a monoclonal anti-CD4 antibody showed that the overexpressed protein was indeed CD4D12. Under reducing conditions, a single band corresponding to CD4D12 was observed (**Figure 21**, a) lanes B-D; b) lane A). However, under non-reducing conditions (**Figure 21**, a) lanes E-G; b) lane B), some binding was also detected at almost twice the mass of the protein, which corresponds to the bands observed by SDS-PAGE.

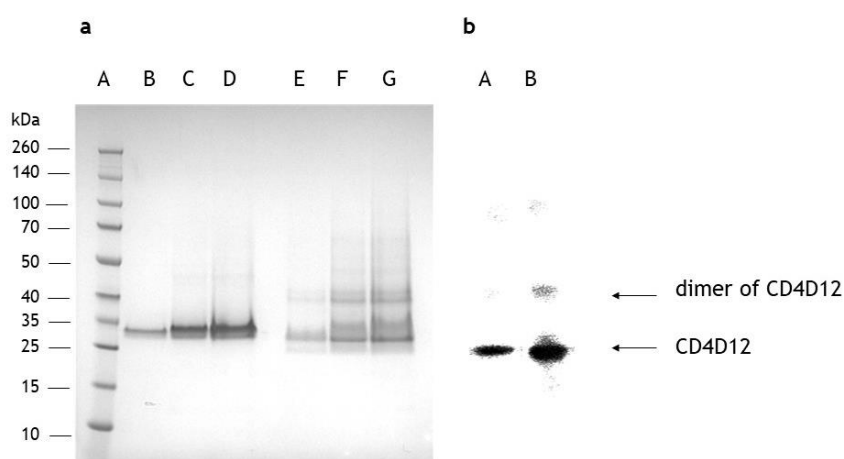


Figure 21 a) SDS-PAGE and b) Western Blot analysis of the purified CD4D12.

a) Lane A: molecular weight marker with values in kDa, Lanes B-D: CD4D12 under reducing conditions at 1.0 µg, 2.5 µg and 4.0 µg in a sample, respectively. Lanes E-G: CD4D12 under non-reducing conditions at 1.0 µg, 2.5 µg and 4.0 µg in a sample, respectively. b) Lane A: CD4D12 under reducing-conditions, lane B: CD4D12 under non-reducing conditions.

These results indicated, that the protein under non-reducing conditions formed intermolecular disulphide bridges. In the overall sequence of CD4D12, there are four cysteine residues and in the native form two disulphide bridges should be

formed: one in the first domain D1, and one in the second domain D2 (**Figure 22**). Since the upper bands were visible both in the SDS-PAGE and in the Western Blot, it suggested that at least a fraction of the entire protein might have formed intermolecular disulphide bonds rather than form into its native conformation. The upper bands were less intense compared to the bands corresponding to the CD4D12.



Figure 22 Crystal structure of CD4D12 (PDB entry: 1GC1). The overall structure of the protein is shown in red. The two disulphide bridges present in the first domain D1 (lower part), and in the second domain D2 (upper part) are shown in yellow.

The next method to confirm the identity of CD4D12 performed with the assistance of Dr. Emma Carrick (Institute of Cardiovascular & Medical Sciences, University of Glasgow) was the in-gel trypsin digest followed by protein mass spectrometry. In the protein mass spectrometry, the protein was initially analysed by SDS-PAGE, the band corresponding to the protein of interest was separated and digested, and the resulting sample was analysed. The peptide sequences were then compared with all human proteins in the UniProt database. Five unique peptide sequences that corresponded to the sequence of the entire 4-domain CD4 protein (UniProt, P01730) were identified:

- NSNQIKILGNQGSFLTK
- KGDTVELTCTASQK
- IDIVVLAFAQ

- ILGNQGSFLTK

- KSIQFHWK

These peptides covered 10.70% of the full length CD4 (Figure 23, also see Appendix Figure 97). All these peptides correspond to regions of the full-length protein that are also present in CD4D12. Thus, the overexpressed protein was CD4D12.

MNRGVPFRHL	LLVLQLALLP	AATQGKKVVL	GKKKGDVELT	CTASQKKSIIQ
FHWKNSNQIK	ILGNQGSFL	TKGPSKLNDR	ADSRRLWDQ	GNFPLIKNL
KIEDSDTYIC	EVEDQKEEVQ	LLVFGLTANS	DTHLLQGQSL	TLTLESPPGS
SPSVQCRSPR	GKNIQGGKTL	SVSQLELQDS	GTWTCTVLQN	QKKVEFKIDI
VVLAFOKASS	IVYKKEGEQV	EFSPLAFTV	EKLTGSGELW	WQAERASSSK
SWITFDLKNK	EVSVKRVTQD	PKLQMGKKLP	LHLTLQALP	QYAGSGNLTL
ALEAKTGKLH	QEVNLVVMRA	TQLQKNLTCE	VWGPTSPKLM	LSLKLENKEA
KVSKREKAVW	VLNPEAGMWQ	CLLSDSGQVL	LESNIKVLPT	WSTPVQPMAL
IVLGGVAGLL	LFIGLGIFFC	VRCRHHRRQA	ERMSQIKRLL	SEKKTCCQCPH
RFQKTCSPI				

Figure 23 Amino acid sequence of the full length CD4 protein (UniProt: #P01730). 5 unique peptide sequences were found within the sequence of the entire CD4 protein. These sequences are highlighted by different colours, moreover one sequence found within a longer peptide is written in red. The underlined sequence corresponds to CD4D12 part of the entire CD4 protein.

Samples of CD4D12 coming from three different purified batches were analysed by CD-spectrometry (see Introduction, section 1.3.3) with the assistance of Dr. Sharon Kelly (Institute of Molecular, Cell and System Biology, University of Glasgow). Samples were at different concentrations: sample 1: 0.505 mg·mL<sup>-1</sup>, sample 2: 0.237 mg·mL<sup>-1</sup>, and sample 3: 2.855 mg·mL<sup>-1</sup> in a PBS buffer at pH 7.5. Far UV CD measurements were measured in 0.02 cm pathlength cuvette. Presence of chloride ions present in samples 1 and 2 precluded measurements below 195 nm for far UV CD spectra. Sample 3 was however 6-fold diluted into H<sub>2</sub>O (final concentration 0.476 mg·mL<sup>-1</sup>), which allowed to obtain data below 195 nm.

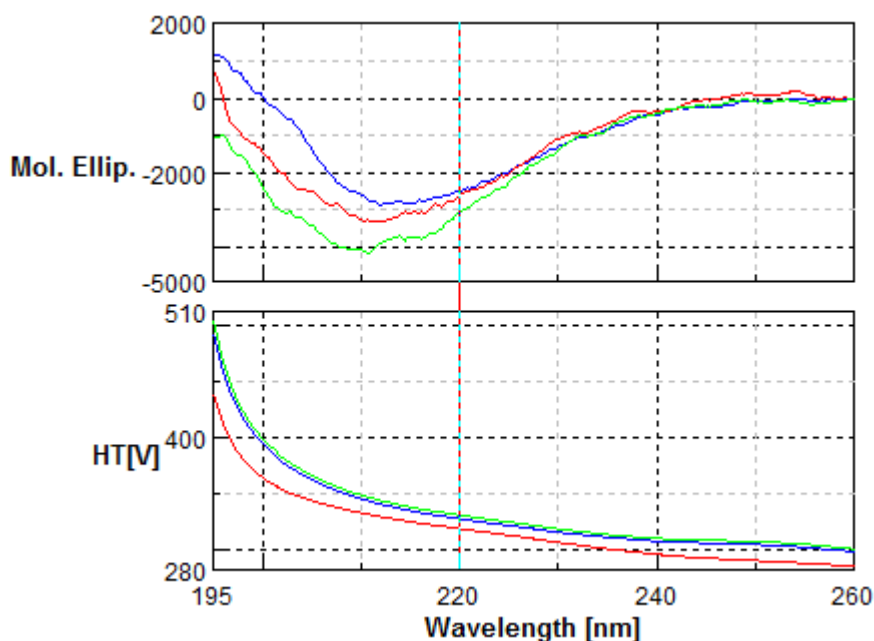


Figure 24 Far UV CD spectra of expressed CD4D12. Sample 1 (blue), sample 2 (red) and sample 3 (green). All spectra have been corrected for protein concentration and cell pathlength.

Negative minima were found to be at different wavelengths in each spectrum: sample 1: ~215 nm, sample 2: ~212 nm, sample 3: ~210 nm (**Figure 24**). These shifts correspond to the shifts found in the literature of ellipticity minima, which were at 212-214 nm for CD4D12.<sup>[115][130]</sup> However, the differences between the shifts may reflect conformational differences between the samples, differences in oxidation states or the presence of contaminating proteins. This would be in accordance with the bands migration on SDS-PAGE and Western Blot.

In the near UV CD measurements peaks in the range of 280-305 nm are characteristic for aromatic amino acid side chains (**Figure 25, a**). Aromatic amino acids: tyrosine (Tyr 99), tryptophan (Trp 45, 79, 174) and phenylalanine (Phe 43, 60, 84, 115, 187, 196) are relatively far away from each other in the primary amino acid sequence (**Figure 25, b, c**). Based on the results from the near UV CD measurements, there was spectral evidence that at least some of the aromatic residues were held rigidly in an asymmetric environment, which indicated that the protein is folded in the region around these residues (**Figure 25**).

Secondary structure estimates were obtained for sample 3, since data could be collected down to 190 nm. The protein was estimated to contain around 35%  $\beta$ -sheet, ~23% turns, ~33% unordered structure and less than 10% helical structure (**Table 4**). These data indicated, that one third of the sample had an unordered

structure, which might be coming from the previously mentioned dimers of CD4D12, which were not in their native conformation. Disulphide bridges formed randomly with other CD4D12 molecules prevented the protein from obtaining its native, correct conformation, therefore the folding process ceased, and some parts of the protein were misfolded. This demonstrated that the refolding procedure was crucial to obtaining a correctly folded protein.

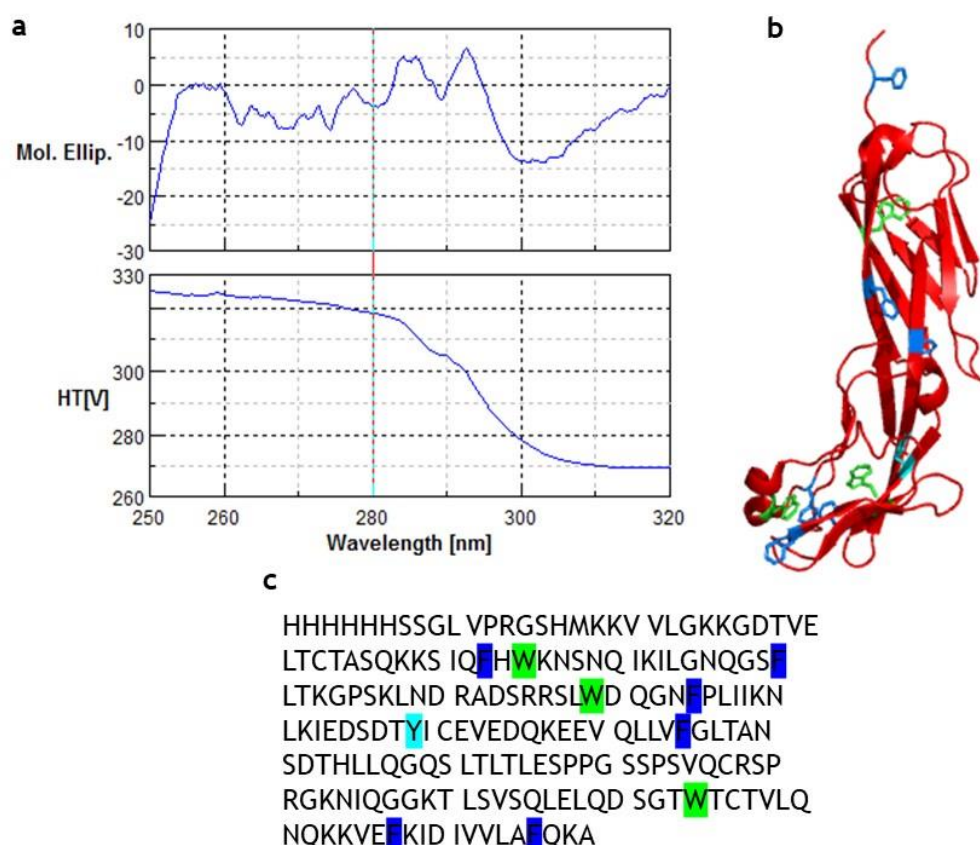


Figure 25 Analysis of secondary structure of expressed CD4D12.

a) Near UV CD spectrum of CD4D12 sample 3, b) crystal structure of CD4D12 (red, PDB entry: 1GC1) with shown aromatic residues: green - Trp, blue – Phe, cyan – Tyr, c) amino acid sequence of the overexpressed CD4D12 with highlighted aromatic residues as in b.

Table 4 Secondary structure estimates – closest matching solution with all proteins. The goodness of fits as judged by NRMSD value (0.055) was favourable. Method used: Contin-LL (Provencher & Glockner method).<sup>[131]</sup>

Helix	Strand	Turns	Unordered	Total
0.096	0.338	0.230	0.336	1

1D <sup>1</sup>H NMR (see Introduction, section 1.3.3) of CD4D12 measurements were performed with the assistance of Dr. Brian Smith (Institute of Molecular, Cell and Systems Biology, University of Glasgow). Although 1D NMR alone cannot be used

for the identification of the protein, it allowed to draw conclusions of how well folded the protein was in solution. In 1D  $^1\text{H}$  NMR spectrum several regions indicated whether the protein was ordered or in a random coil form.<sup>[132][133]</sup> Well dispersed signals in the aliphatic region (-1.0-1.0 ppm) indicate a folded protein, in contrast to a sharp signal at approximately 1 ppm, indicating an unfolded protein. Large, broad peaks in the region of 8.3 ppm are a fingerprint of the amide backbone in a random coil configuration. Conversely, well dispersed peaks in this region indicate a folded protein. Another indication of a folded protein are dispersed peaks in the region of 8.5-11 ppm.

The purified protein sample in PBS buffer, pH 7.5, was supplemented with  $\text{D}_2\text{O}$  to a final concentration of 5% and the 1D NMR spectrum was recorded (**Figure 26**). Well dispersed peaks especially in the  $\text{CH}_3$  region (-0.5-1.0 ppm) as well as peaks present downfield, characteristic to  $\text{CH}_2$  and  $\alpha$  C-H protons in the region between 1-4.5 ppm, indicate hydrophobic interactions present in the proteins that has a rigid secondary structure. The region between 6.5-8.5 ppm corresponds to aromatic CHs as well as backbone amide protons. This region is well dispersed, although three characteristic peaks at 6.77, 7.65 and 8.37 ppm are distinct. These are possibly protons of the imidazole ring present in histidines. The overexpressed CD4D12 protein has nine histidines and six out of them are within the His-tag at the N-terminus. It was also observed, that the protons corresponding to the indol ring in tryptophans are not visible. This might be caused by rapid exchange of protons with the solvent when the indol protons are not protected by other residues.

Several broader peaks are also visible in the spectrum at 1.1, 1.60 and 2.92 ppm. The overall pattern of the 1D  $^1\text{H}$  NMR indicates, that the protein is a mixture of both properly folded protein with a smaller fraction of misfolded protein.

One of the possibilities to confirm the folding pattern of the protein would be to record additionally the 1D spectrum of this protein in the presence of denaturant. In this way it would be possible to compare the spectra of unfolded and folded protein.

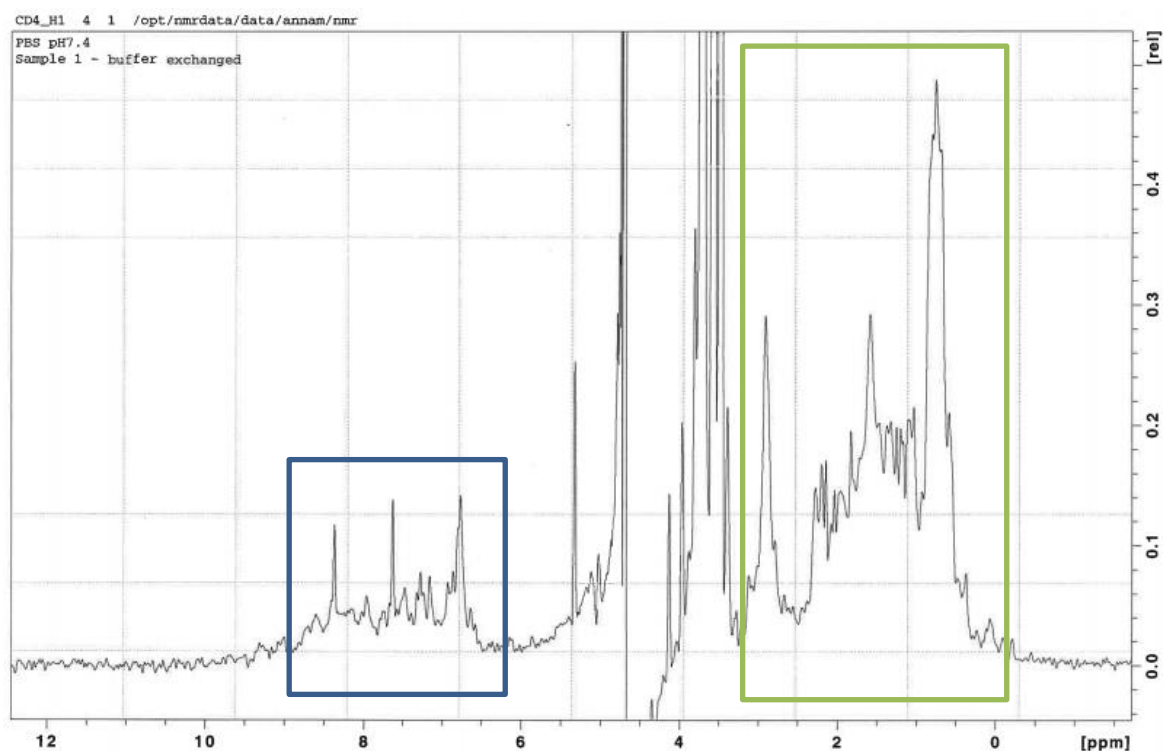


Figure 26 1D  $^1\text{H}$  NMR of CD4D12 protein.

Scale in ppm. The water signal at 4.5 ppm was suppressed. Blue box indicates imidazole ring present in histidines. Green box indicates well dispersed peaks of the aliphatic region.

Another method to confirm whether the protein was in the correctly folded form, which is biologically active, is to evaluate its binding properties towards a known ligand and to compare the obtained values with the literature data. This has been described in the chapter 4.2.1 in which evaluation of the CD4D12 activity was performed by SPR.

## **2.2 Expression of HIV-gp120 in mammalian cells**

### **2.2.1 HIV-gp120**

Gp120 protein was required as a positive control for binding experiments, thus efforts towards its expression and purification were undertaken.

The structure of gp120 and the mechanism of binding to CD4 followed by viral entry of HIV into the cell have been described in Chapter 1.

### **2.2.2 Overview of protein production in mammalian cells**

As previously discussed in section 2.1.2 protein overexpression in eukaryotic cells has the advantage of production of biologically active proteins folded into their native form and equipped with post-translational modifications, such as glycosylation or phosphorylation.<sup>[120]</sup> Protein expression can be performed in a variety of different cell lines, among the most frequently used ones: HeLa, Baby Hamster Kidney (BHK) cells, Human Embryonic Kidney (HEK293) cells or Chinese Hamster Ovary (CHO) cells. The most common applications of the mammalian expression system include: expression of complex proteins, virus and antibody production, functional assays. However, this expression system is also highly challenging due to low yields in protein production by adherent cells, demanding culture conditions, high costs of tissue culture and potential contaminations of the cells.<sup>[134]</sup>

Gp120 is a highly glycosylated protein and glycans are required for the correct folding of the protein,<sup>[13][14]</sup> therefore HEK293 cells were chosen as the expression system, which was described previously for the obtained plasmid.<sup>[128]</sup>

### **2.2.3 Overview of gene transfection**

Transfection is the introduction of the foreign genetic information into eukaryotic cells in order to change properties of the cell to either produce recombinant protein of interest or inhibit gene expression. Transfection can be broadly classified into two general types: transient and stable transfection. In transient transfection the introduced genetic information is not integrated into the genome of the host cell and therefore is present in the cell only for a limited period of



time. Dividing cells do not pass the introduced genetic material to the next generations the introduced genetic material. The advantage of this approach is that the material is transfected in a high copy number which results in high expression levels of the protein. In the stable expression strategy, the introduced genetic material is integrated into the host genome, which allows its passage to the next generations during cell division. On the other hand, the number of copies of the transfected material is kept low, which results in lower yields of expressed protein when compared to transiently transfected genes.<sup>[135]</sup>

The cell membrane carries negative charge on its outer surface which prevents entry of similarly charged nucleic acids into the cell. Therefore, chemical, biological and physical methods were developed for gene delivery to the cell. Chemical methods are based on neutralisation of the negative charge of the nucleic acid by carrier molecules. Biological methods employ genetically engineered viruses to transfer the gene of interest into the cell. Physical methods aim at creation of transient pores in the cell membrane, for transfer of the nucleic acid to pass into the cell.

Once the genetic material is transfected into the eukaryotic cell, after integration in cellular genome, it will be transcribed in nucleus to produce messenger RNA (mRNA) and translated to give the desired protein which will be post-translationally modified to produce a functional protein.<sup>[136]</sup>

## **2.2.4 Results and discussion**

### **2.2.4.1 Restriction enzyme digestion of the V1Jnstpagp120 plasmid**

Our group has contacted professor Varadarajan (Indian Institute of Science, Bangalore, India) with respect to the plasmid used for expression of the full length gp120. DNA plasmid V1Jnstpagp120<sup>[128]</sup> was kindly donated to our group.

Initially, the donated plasmid was analysed by restriction enzyme digestion. The V1Jnstpagp120 plasmid consists of 6391 base pairs (**Figure 27**) and possesses restriction sites: SspI and BlnI. Restriction enzyme digestion was performed as described in the Materials and Methods section 6 and samples prior and post digestion were analysed by DNA electrophoresis. Prior to the restriction digest (lane C) the plasmid migrated above 5,000 bp. Restriction enzyme digestion of the

plasmid with enzymes SspI and BlnI produced two DNA fragments of 1818 bp and 4573 bp, as predicted (lane B). This result confirmed the size of the plasmid.

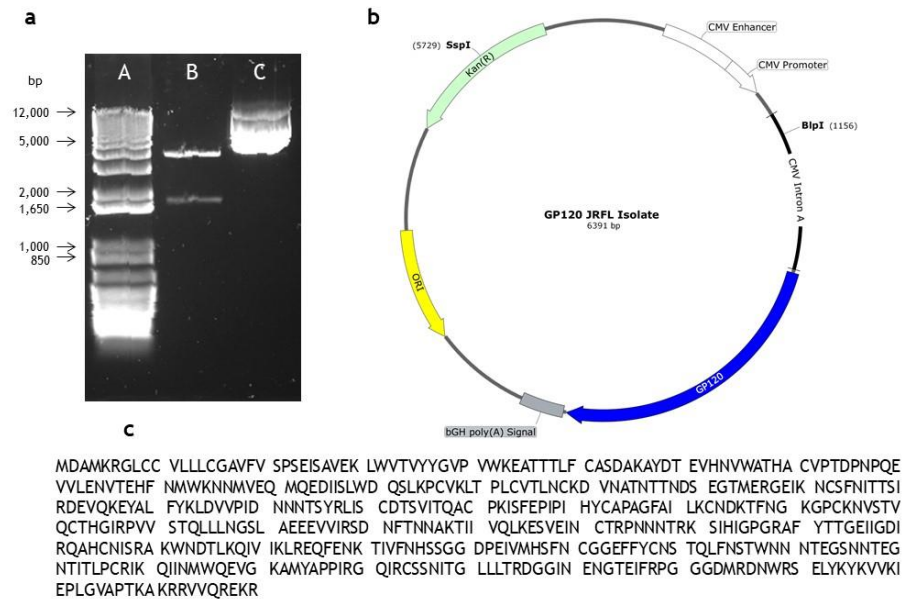


Figure 27 a) DNA electrophoresis of restriction enzyme digest of V1Jnstpagp120 plasmid. Lane A: 1 Kb Plus DNA ladder, lane B - the cut plasmid after restriction enzyme digestion, lane C - the plasmid prior to restriction digest. b) Vector map of V1Jnstpagp120 plasmid, c) the amino acid sequence of gp120.

#### 2.2.4.2 Plasmid transfection and protein expression

The transfection method was altered from the original method described in the literature by Varadarajan et al.<sup>[128]</sup> A polymer transfection reagent jetPRIME™ reagent was used in order to deliver the plasmid into the cells.

Test expression and purification of gp120 showed several challenges. Firstly, SDS-PAGE of all obtained samples was performed (**Figure 28**). It revealed that three different populations of gp120 were expressed: nonglycosylated monomer, glycosylated monomer and glycosylated dimer. As can be observed, quite a large portion of the expressed protein was expressed in the nonglycosylated form. Also, the band corresponding to the dimer of gp120 seemed more distinct than its monomeric form. Secondly, binding conditions during the purification should be optimised, because a large portion of proteins was eluted in the flow-through. The bands corresponding to the glycosylated monomer of the gp120 were not migrating exactly at 120 kDa, which could have been caused by heterogeneity in protein glycosylation.<sup>[137]</sup>

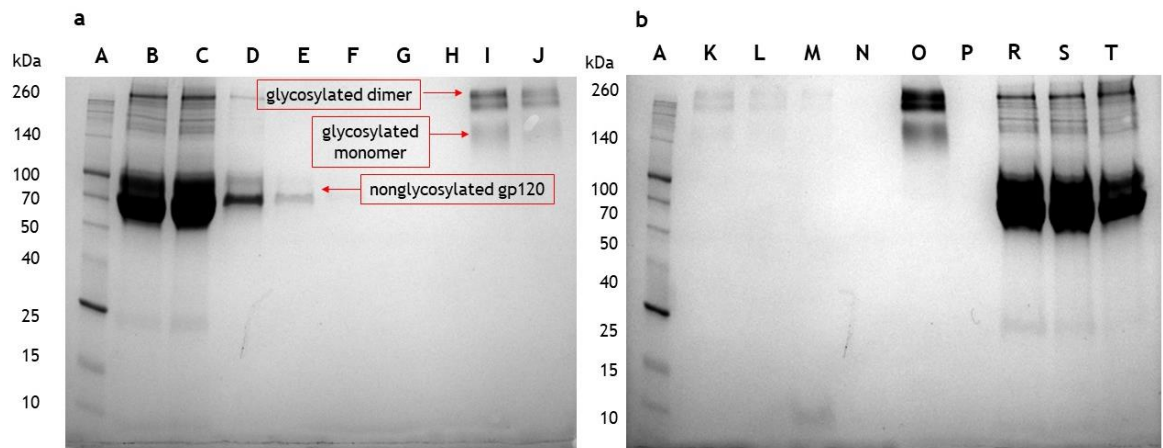


Figure 28 SDS-PAGE analysis of gp120 expression in HEK293 cells.

Lanes A: molecular weight marker with values in kDa, lane B: tissue culture supernatant after 24 h post-transfection, lane C: flowthrough, lanes D-F: washes with wash buffer, lanes G-H: washes with binding buffer, lane I-L: elution fractions of the gp120 protein, lane M: resin beads post-purification, lane N: filtrate post-concentration, lane O: concentrated eluted fractions, lane P: empty, lane R: tissue culture supernatant after 48 h post-transfection, lane S: tissue culture supernatant after 72 h post-transfection, lane T: tissue culture supernatant after 72 h post-transfection in Optimem-reduced medium.

The difference in expression level of gp120 can be compared by analysis of bands corresponding to the tissue culture supernatant coming from the samples 24, 48, and 72 h after transfection (**Figure 28**, lanes B, R, S, T). The expression time seemed to be insignificant, however, it was later shown by Western Blot analysis with anti-gp120 antibody, that the nonglycosylated form of the gp120 was formed mostly inside the cell and was not secreted into the tissue supernatant. The amount of glycosylated form varied especially between 24 h and 48 h after induction. It could be observed that the longer the plasmid was incubated with the cells, the more of the glycosylated protein was produced. There was a small increase in gp120 production after 72 h, however when the cells were incubated in serum reduced medium, the amount of the produced protein decreased (**Figure 29**). The observation that gp120 was secreted into the tissue culture supernatant is in accordance with the fact, that gp120 contains in the sequence a signal peptide (residues 1-31), which allows the protein to be secreted into the tissue culture supernatant.<sup>[138]</sup> From thereafter, the cells were incubated for 72 h post-transfection.

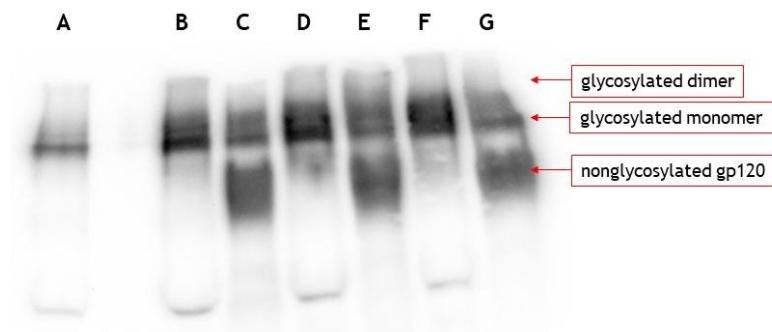


Figure 29 Western Blot analysis of the gp120 expression post-induction after 24, 48, 72 h. Lane A: tissue supernatant after 24 h, lane B: tissue supernatant after 48 h, lane C: cell lysate after 48 h, lane D: tissue supernatant after 72 h, lane E: cell lysate after 72 h, lane F: tissue supernatant after 72 h in serum reduced medium, lane G: cell lysate after 72 h in serum reduced medium.

### 2.2.4.3 Overview of gp120 purification

As was described in the Chapter 1, gp120 is a heavily glycosylated glycoprotein present on the surface of the HIV virus. *N*-linked glycans contribute to approximately half of the molecular weight of the entire glycoprotein. The majority of the glycans are mannoses: Man-5 and Man-9.<sup>[13][14]</sup> This structural feature of the gp120 structure can be used during the protein purification because sugar residues present in the glycoproteins can bind to carbohydrate-binding proteins, that is lectins. Lectins are present on the stationary phase of the column and bind glycoproteins selectively, enabling their isolation from a cell lysate. Once the impurities are washed away, the glycoprotein is eluted with methyl- $\alpha$ -D-mannopyranoside (MMP). Detergent is added to the solubilisation, wash and elution buffers to reduce non-specific binding to the column.

#### 2.2.4.3.1 Purification of gp120

Purification was performed either by batch method using *Galanthus nivalis* (GNL) lectin-bound agarose beads or by column packed with Lentil Lectin resin, followed by size exclusion chromatography. Purification of the collected supernatant after transfection was performed according to a slightly modified literature procedure.<sup>[139]</sup> The first step of the purification which employed lectin beads was used to isolate glycoproteins present in the tissue culture supernatant from all the other proteins produced by the cells. This step yielded relatively pure protein, however it was observed that the monomer and dimer species eluted almost together, as the resin does not distinguish between them (Figure 30 a, b). Size exclusion chromatography allowed to obtain highly pure protein separated from dimeric species (Figure 30, c, lane K), but the purified gp120 was obtained in low

yield and was highly unstable, as it precipitated out of the solution. In the future, it would be worth considering adding another step to the purification, such as dialysis, to remove detergent from the protein sample prior to size exclusion chromatography. ITC measurements for binding studies of discontinuous mimics required large volumes of highly concentrated gp120 as a positive control. To obtain gp120 in these quantities, the expression would have to be immensely scaled up and optimised, which would be extremely time-consuming. Therefore, work on gp120 expression and purification was discontinued.

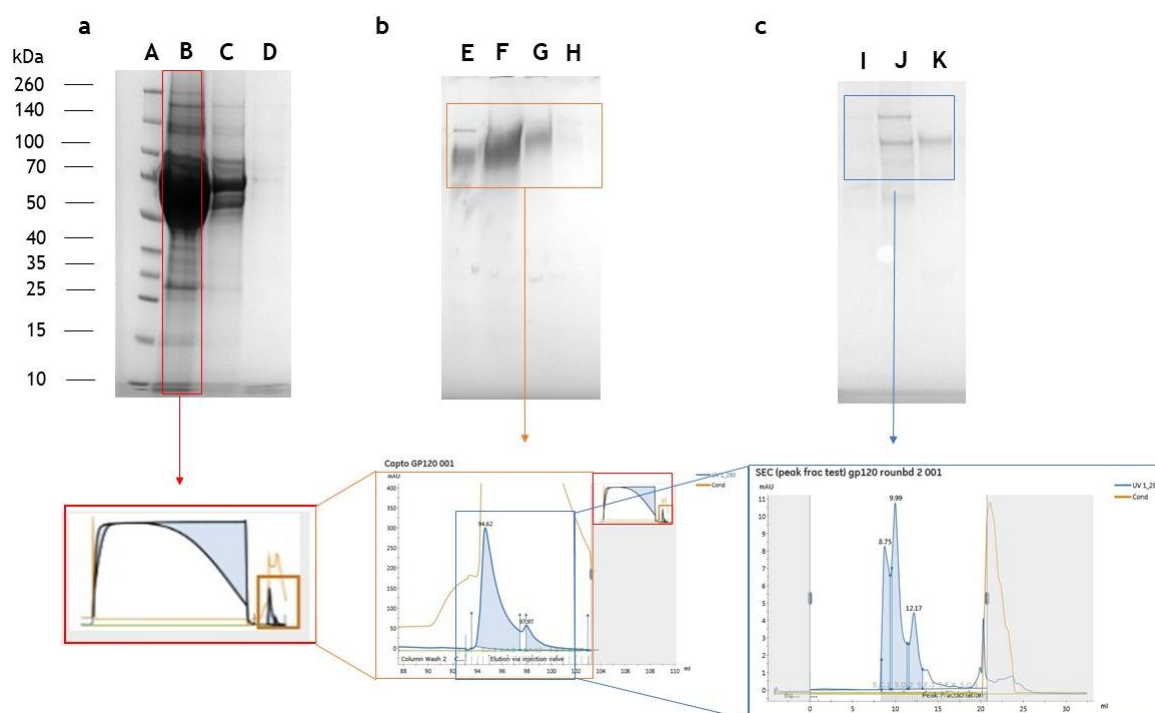


Figure 30 Progress of purification of gp120 by SDS-PAGE and UV.

a) and b) upper: SDS-PAGE analysis of the first step purification of gp120 by Lentil-Lectin column: Lane A: molecular weight marker with values in kDa, Lane B: flowthrough, Lane C-D washes, Lane E-H: eluted fractions of gp120; lower: chromatograms of the performed purification. c) upper SDS-PAGE analysis after size exclusion chromatography. Lanes I-K: eluted fractions of gp120; lower: chromatogram of the performed purification.

#### 2.2.4.4 Confirmation of identity of gp120

As was discussed in section 2.2.4.2, the Western Blot analysis performed with a polyclonal anti-gp120 antibody identified that the overexpressed protein was indeed gp120. However, it was observed that the antibody bound also the nonglycosylated form of the glycoprotein.

Samples of gp120 coming from two different purified batches were analysed by CD-spectrometry with the assistance of Dr. Sharon Kelly (Institute of Molecular, Cell and Systems Biology, University of Glasgow). Samples were at different

concentrations: sample 1:  $267 \mu\text{g}\cdot\text{mL}^{-1}$  and sample 2:  $296 \mu\text{g}\cdot\text{mL}^{-1}$  in a PBS buffer at pH 7.5. Presence of chloride ions present in samples 1 and 2 precluded measurements below 195 nm for far UV CD spectra, therefore only far UV spectra were recorded. The spectra appeared to be very similar in intensity and shape and suggested that the protein in each fraction was folded (**Figure 31**). Negative minima were found at approximately 210 nm and the overall shape of the spectrum resembled closely the CD spectra found in the literature.<sup>[128][140]</sup>

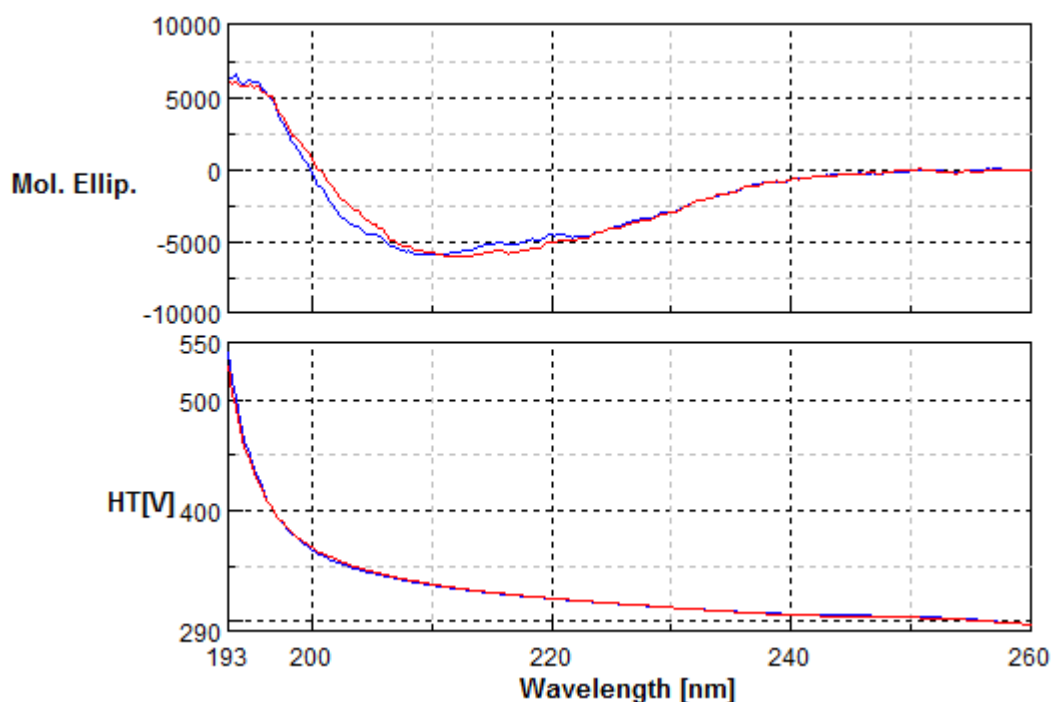


Figure 31 Far UV spectra of gp120 samples: sample 1 (blue), sample 2 (red).

Another method to confirm whether the protein is in the correctly folded form, which is biologically active, is to evaluate its binding properties towards a known ligand and to compare the obtained values with the literature data. An ITC experiment was performed between the expressed, purified gp120 and expressed and purified CD4D12. However, the concentrations of the samples were by far too low to detect any binding, therefore it could not be concluded whether gp120 was expressed in the biologically active form.

## 2.3 Conclusions

Efforts towards expression, purification and characterisation of CD4D12 and gp120 proteins were made.

CD4D12 was a challenging protein to express in *E.coli*, refold and purify. It could be produced on a large scale and the scale-up was feasible. It was characterised by various methods including: SDS-PAGE, Western Blot, 1D <sup>1</sup>H-NMR and CD-spectroscopy, and protein mass spectroscopy. All these techniques confirmed that CD4D12 protein was expressed and the majority of it was properly folded. However, finding refolding conditions to obtain correctly folded, biologically active protein was an extremely challenging task. Several refolding methods were tested, yet optimisation is still required in order to avoid formation of the improper dimeric species coming probably from the incorrectly formed intermolecular disulphide bridges.

The glycoprotein gp120 was transiently expressed in mammalian cells, which was confirmed by SDS-PAGE and Western Blot. Gp120 requires at least a two-step purification, and future wise it would benefit from introduction of an intermediate step to remove the detergent from the sample. Since the expression in the current setup is rather low yielding, a scale up of the procedure would be required. A possible remedy would be to employ Vaccinia Virus as a tool for gene expression in mammalian cells<sup>[139]</sup> or protein in baculovirus-infected insect cells.<sup>[141]</sup>

ITC studies planned for evaluation of binding properties would require initial optimisation of the experimental conditions, which only then could be followed by the actual experiments. Low-yielding expression of gp120 and requirement of large volume of highly concentrated sample for the ITC studies resulted in discontinuation of gp120 expression. Thus, it was decided, that the binding studies should be evaluated by SPR, which consumes less proteins.

## 3 Synthesis of the gp120 discontinuous epitope mimics

### 3.1 Introduction

After 30 years of major global efforts towards the discovery of a HIV-1 vaccine, a successful candidate capable of prevention of the viral infection has thusfar not been identified. Research in this field is an intriguing challenge. HIV-1 and its interactions with the CD4 receptor and CCR5 and CXR4 co-receptors, as well as its impact on the immune system are very well documented and studied. The abundance of the literature on the topic indicates that plentiful ideas and scientific approaches have already been investigated, however, without success. Breakthroughs in this area require exceptionally creative and innovative approaches.

Previous research carried out in our group towards the design and synthesis of a HIV-1 vaccine served as a good starting point for further optimisation and evaluation of the constructs.<sup>[142][143][144][145]</sup> The TAC scaffold developed in our group was chosen as a platform for attachment of the cyclic peptides because of the well-developed synthetic route and the ease of preparation of the scaffold on a large scale. Moreover, Werkhoven et al.<sup>[88]</sup> have shown that orthogonally protected alkynes present on the TAC scaffold allowed the introduction of three different cyclic peptides bearing an azide handle with a control over positioning of the loops. Introduction of the cyclic peptides onto such scaffold allowed to display them in a pre-organised fashion mimicking the discontinuous site of gp120. The peptide sequences selected for the mimicry of gp120 were based on the established conserved amino acids corresponding to the CD4-binding site found in gp120.<sup>[16]</sup> These are: loop 1: <sup>454</sup>LTRDGGK<sup>460</sup>, loop 2: <sup>424</sup>INMWQEVGKA<sup>433</sup> and loop 3: <sup>365</sup>SGGDPEIVT<sup>373</sup> (**Figure 32**). The cyclisation of the peptides was performed to improve their stability as well as induce a loop conformation, which should improve the mimicry of the naturally occurring conformation in gp120 when compared to the linear peptides. Moreover, Meuleman et al.<sup>[146]</sup> have recently shown, that cyclic peptides mimic epitopes more closely than their linear counterparts and have improved specificity.



Introduction of the peptide loops was performed by copper catalysed azide-alkyne cycloaddition (CuAAC, also click). The CuAAC method of ligation is a stereoselective reaction between a terminal azide and alkyne moieties catalysed by copper (I) species. The product of this reaction is a 1,4-disubstituted 1,2,3-triazole. The versatility of its applications is the main cause why the click reaction is frequently employed for conjugation of peptides and biomolecules. The reaction can be performed in various solvents and is insensitive to pH and temperature. CuAAC is well documented to be bio-compatible, efficient, robust, and fast. Moreover, it allows the introduction of unprotected peptides. For these reasons CuAAC was chosen as the method of introduction of cyclic peptides representing discontinuous epitopes of gp120 onto the TAC scaffold, which orthogonally protected alkyne handles are compatible with CuAAC chemistry.<sup>[147][148]</sup>

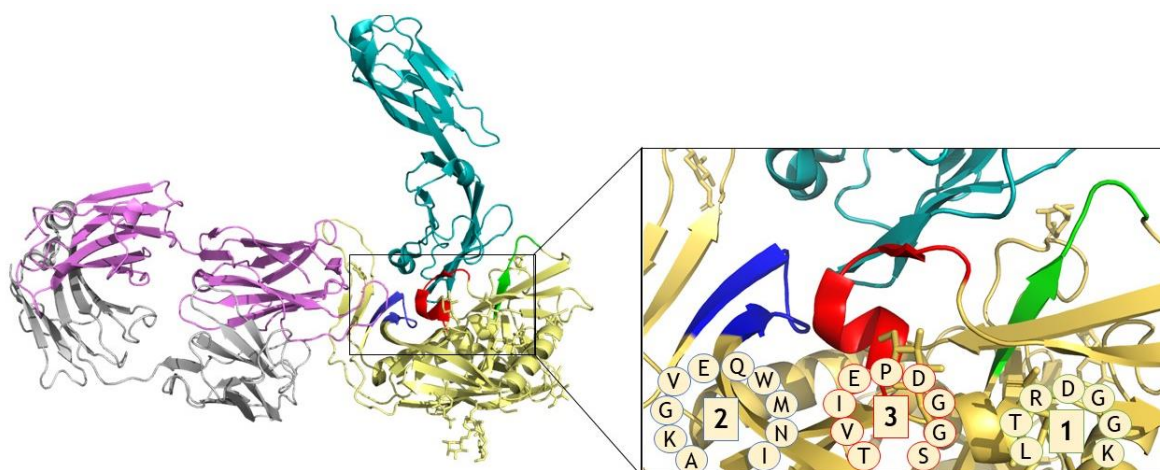


Figure 32 X-ray crystal structure of gp120-CD4-antibody complex (PDB entry: 1GC1). Left: gp120 (yellow) binds CD4 (cyan) via the loops present in the conserved CD4-binding site (blue, red and green). The antibody (grey and pink) interacts with the variable V1/V2 region of gp120, which is more accessible than the CD4-binding site. Right: zoom of the conserved residues within CD4-binding site of gp120.

The TAC molecular scaffold should position the attached discontinuous epitope sequences in the correct spatial conformation, thereby mimicking the gp120 CD4-binding site. Utilisation of the scaffold allowed omitting the bulk of the entire protein and possibly avoid the viral immune evasion mechanisms, such as glycan shielding or presentation of the immunodominant epitopes (**Figure 33**). It is known that the CD4-binding site is conserved among different HIV-1 strains.<sup>[149]</sup> Therefore, a mimic of this conserved region in a vaccine would not only be broadly applicable to different HIV-1 strains but would also be resistant against emergence

of new viral mutants. A synthetic mimic of the gp120 discontinuous epitopes synthesized in this way might prove to be an efficacious HIV-1 vaccine candidate.

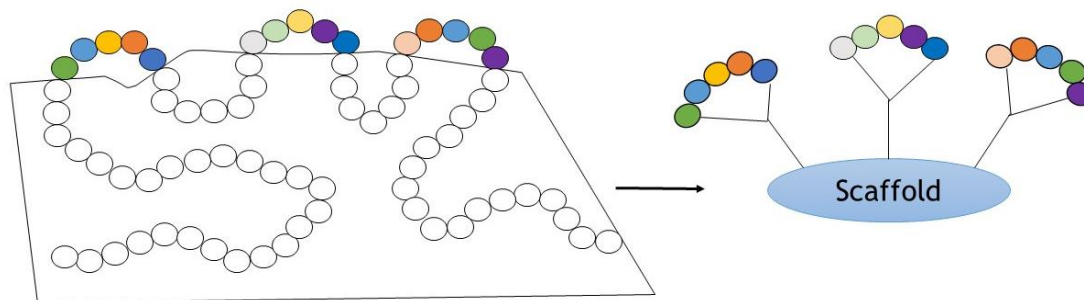


Figure 33 Attachment of the cyclic peptides onto the scaffold for the mimicry of the discontinuous epitopes.

Peptide sequences analogous to the protein discontinuous epitopes are synthesized and cyclised, which is followed by their attachment on to the molecular scaffold. The scaffold serves as a skeleton positioning the loops in the right spatial conformation and omits the bulk of the protein.

The main aim of this project was to synthesize the discontinuous epitope mimics of the HIV-1 gp120 by attachment of the corresponding cyclic peptides to the TAC scaffold, which is described in this chapter. The process of the synthesis can be separated into three parts: synthesis of the TAC scaffold, synthesis of the cyclic peptides corresponding to the CD4-binding site on gp120 and lastly, attachment of the peptidic loops onto the scaffold by the means of CuAAC chemistry.

Moreover, first attempts of characterisation by NMR and CD-spectroscopy were performed on the linear and cyclic peptides as well as the scaffolded cyclic peptides to gain further insight into their structural features. The combination of these two techniques proved to be a powerful source of information on the three-dimensional conformation of the studied molecules. Each compound was studied separately to analyse whether it obtained a certain secondary structure or if it was disordered. Afterwards, NMR spectra of the cyclic peptides were superimposed on the spectrum of the entire constructs to analyse the shift differences. This method allowed to draw conclusions on the structural organisation of the molecules and could aid optimisation of their design to obtain biologically active compounds.

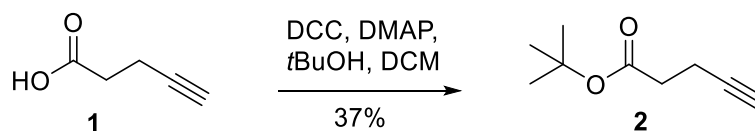
Synthesized constructs and cyclic peptide loops described below were tested for their biological activity, which will be described in section 4.2.2.

## 3.2 Results and discussion

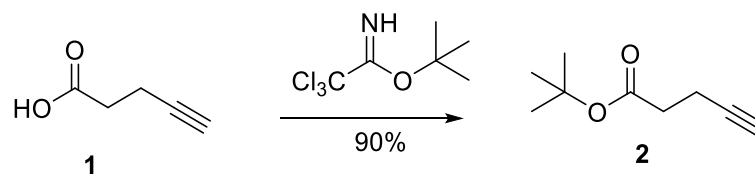
### 3.2.1 Synthesis of TriAzaCyclophane (TAC) scaffold

Synthesis of the TAC scaffold was performed according to our slightly modified literature procedure.<sup>[88]</sup> The TAC scaffold was equipped with two protected alkynes, which are triethylsilyl (TES) and triisopropylsilyl (TIPS) derivatives of pentynoic acid. These compounds were synthesized in three steps. Firstly, the carboxylic acid moiety of 4-pentynoic acid was protected by conversion to a *tert*-butyl ester (**Scheme 1**). Initially this reaction was performed according to the literature procedure by Fischer et al,<sup>[150]</sup> however, the product was obtained consistently in low yields (37%). *Tert*-butyl-protected pentynoic acid was required in large amounts as it was a precursor to obtain TES and TIPS-protected pentynoic acid derivatives. To increase the yields, a new method was investigated.<sup>[151]</sup> 4-pentynoic acid was reacted with *tert*-butyl-2,2,2-trichloroacetimidate to afford the desired product in an excellent yield of 90%. Moreover, this approach benefited from an easy work-up procedure, which was a simple filtration using Celite to remove the crystalline side product trichloroacetamide. Due to volatility of the obtained compound **2**, it was crucial to work with solvents that were easy to evaporate during the work-up (DCM), in following reactions and respective purifications (Et<sub>2</sub>O, pentane). The following reactions of TES and TIPS-protection had the highest yields when performed in THF (described further below), which was present in a small amount. Although it has a higher boiling point than Et<sub>2</sub>O, it could be still evaporated, if care was taken. The purification had to be performed in Et<sub>2</sub>O/pentane. After the purification the product was highly diluted, and it could not be recovered if the purification was performed with ethyl acetate and hexane or petroleum ether. Upon evaporation of these solvents, the product evaporated too.

The original method:



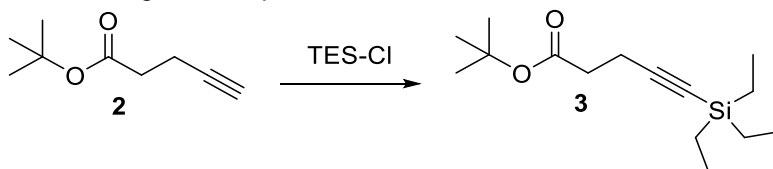
The new method:



Scheme 1 Synthesis of the *tert*-butyl protected pentynoic acid **2**.

The next step involved the reaction of the afforded *tert*-butyl ester **2** with an appropriate silyl chloride to obtain TES and TIPS-protected alkynes. When the reaction was performed according to the published procedure (Table 5, entry 1), the product was obtained in a lower yield (Table 5, entry 2), therefore screening of the reaction conditions to improve the yield was performed. The alterations included changes of the solvent in which the reaction was performed and the number of equivalents (equiv.) of added *n*-BuLi. It was found that an increase in the amount of *n*-BuLi from 1.0 to 1.2 equiv. decreased the yield to 35% (Table 5, entry 3). Change of the solvent to  $\text{Et}_2\text{O}$  was thought to improve the yield of the recovered product after the work-up, since it was found that these compounds were volatile. It was observed that this reaction proceeded with formation of a milky precipitate, which was not observed previously and the reaction yield decreased further down to 13% (Table 5, entry 4). Even though the reported yield could not be reproduced, the initial conditions (Table 5, entry 2) were applied in TES-protection of *tert*-butyl 4 pentynoate **3** and enough of the product was obtained to proceed.

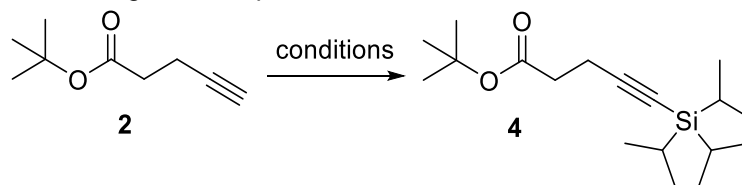
Table 5 Conditions screening for TES-protection.



Entry	Solvent	<i>n</i> BuLi equiv.	Yield
1. <sup>[88]</sup>	THF anh.	1.0	72%
2.	THF anh.	1.0	59%
3.	THF anh.	1.2	35%
4.	Et <sub>2</sub> O anh.	1.0	13%

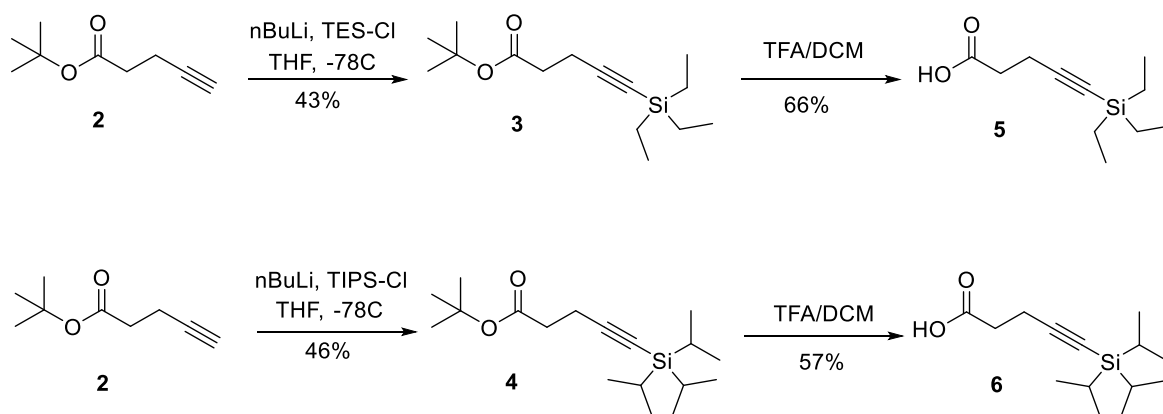
A similar screen of reaction conditions was performed for TIPS-protection of the *tert*-butyl ester. Performance of the reaction under published conditions (Table 6, entry 1) resulted in a lower yield (Table 6, entry 2). Since it was observed, that the starting material was still present at the end of the reaction, the amount of *n*-BuLi was increased up to 1.8 equiv. to increase the deprotonation rate, however, the yields under these conditions decreased significantly (Table 6, entry 3). The impact of dry Et<sub>2</sub>O as a solvent was investigated also in the TIPS-protection reaction. When the reaction was performed under the same conditions as the literature reaction with only the solvent changed to Et<sub>2</sub>O, the yield decreased down to 4% (Table 6, entry 4). However, when LDA was used instead of *n*-BuLi as a base in Et<sub>2</sub>O, the yield increased to 21% (Table 6, entry 5). Moreover, the alteration of TIPS-Cl to TIPS triflate was investigated. This reaction also proceeded with low yields (Table 6, entry 6). Even though the reported yield could not be reproduced, the initial conditions (Table 6, entry 2) were applied in TIPS-protection of *tert*-butyl 4-pentynoate **4** and enough of the product was obtained to proceed.

Table 6 Conditions screening for TIPS-protection.



Entry	Solvent	Base (equiv.)	TIPS source (equiv.)	Yield
1. <sup>[88]</sup>	THF anh.	<i>n</i> -BuLi (1.0)	TIPS-Cl (1.2)	68%
2.	THF. anh.	<i>n</i> -BuLi (1.0)	TIPS-Cl (1.2)	46%
3.	THF anh.	<i>n</i> -BuLi (1.8)	TIPS-Cl (1.2)	6.6%
4.	Et <sub>2</sub> O anh.	<i>n</i> -BuLi (1.0)	TIPS-Cl (1.2)	4%
5.	Et <sub>2</sub> O anh.	LDA (1.0)	TIPS-Cl (1.2)	21%
6.	THF anh.	<i>n</i> -BuLi (1.1)	TIPS triflate (1.0)	16%

In a final step of the preparation of TES and TIPS-protected alkynes, the *tert*-butyl ester was removed by treatment with TFA. In the literature procedure<sup>[88]</sup> *tert*-butyl protected silyl alkynes were treated with 15% TFA/DCM. However, it was found that the amount of TFA could be decreased to 10%. These reactions proceeded with good yields of 66% (for compound **3**) and 57% (for compound **4**). Scheme 2 summarizes the final conditions for the synthesis of the TES and TIPS-protected 4-pentynoic acid derivatives.



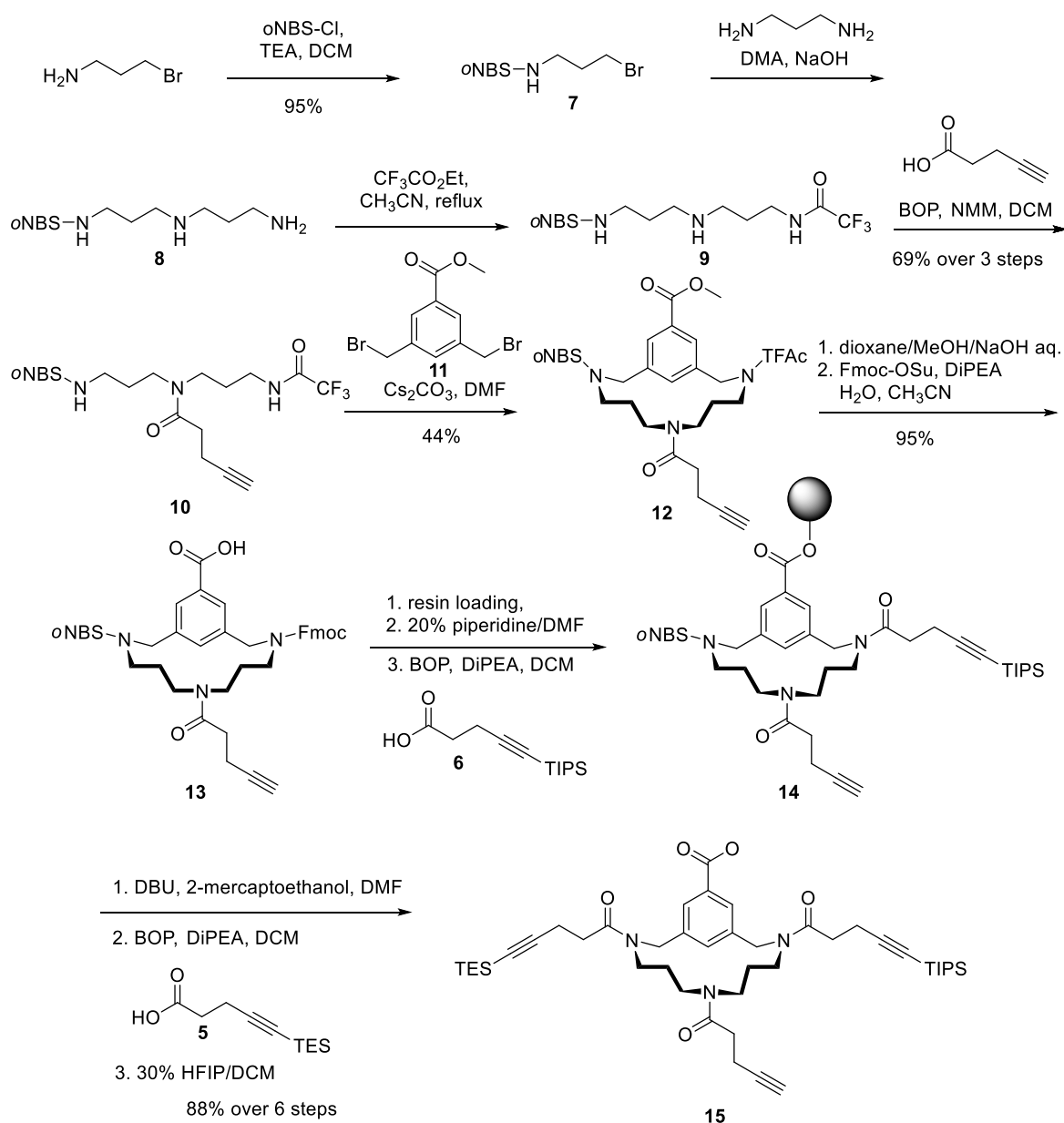
Scheme 2 Synthesis of the TES and TIPS protected pentynoic acid derivatives.

Synthesis of the TAC scaffold was performed according to the literature procedure published recently by Werkhoven et al.<sup>[88]</sup> The synthesis route is shown in the **Scheme 3**. The first step in the synthesis was protection of 3-bromopropylamine with the *o*-nitrobenzenesulfonyl (*o*NBS) protecting group to afford *o*NBS-protected bromopropylamine **7** in an excellent yield of 95%. Next, bromopropylamine **7** was reacted with 10 equiv. of 1,3-diaminopropane in DMA. After stirring overnight, 1 equiv. of 4.0 M NaOH was added to deprotonate the amines, followed by evaporation of the excess diamine, which had a lower boiling point (140°C) than DMA (164.5-166°C). The reaction mixture was co-evaporated with DMA until the collected solvent was no longer basic. The crude product **8** was treated with ethyl trifluoroacetate to obtain bis-protected triamine **9**. 4-pentynoic acid was coupled to the secondary amine functionality of triamine **9** in the presence of BOP and NMM, which after purification yielded product **10** in the yield of 69% over three steps.

The triple-protected triamine **10** was cyclised with the dibromide **11** in presence of Cs<sub>2</sub>CO<sub>3</sub>, to obtain the TAC-scaffold skeleton **12** in a yield 44%. The cyclisation was performed at a relatively high dilution (12 mM) to prevent formation of the dimeric cyclic side product. The next step was the removal of trifluoroacetyl protecting group and methyl-ester by hydrolysis with Tesser's base (15:4:1 dioxane/MeOH/aqueous NaOH 4 M).<sup>[152]</sup> This was followed by Fmoc-protection of the free amine, which afforded the *o*NBS,Fmoc-protected TAC **13** scaffold in an excellent yield of 95%.

The scaffold was loaded onto a 2-chlorotrityl resin and the Fmoc group was removed by treatment with 20% piperidine/DMF. At this point the TIPS and

TES-protected pentynoic acid derivatives were attached to the TAC scaffold one after the other. It is known, that the TIPS-protecting group is more stable than the TES-protecting group, therefore it was decided, that TIPS will be introduced first.<sup>[88]</sup> The TIPS-protected pentynoic acid **6** was coupled to the scaffold using BOP to afford the product **14**. Afterwards, the *o*NBS-group was removed by treatment with  $\beta$ -mercaptoethanol and DBU, which was followed by coupling of the TES-protected 4-pentynoic acid **5** to the scaffold. Cleavage of the scaffold from the resin was accomplished by treatment with HFIP, which afforded the orthogonally protected trialkyne TAC scaffold **15** in the yield of 88%.



Scheme 3 Synthesis of the orthogonally protected trialkyne TAC scaffold.



### 3.2.2 Synthesis of the cyclic peptides

The next step in the preparation of discontinuous epitope mimics was the synthesis of the cyclic peptides corresponding to the conserved CD4-binding site of HIV-1 gp120. The sequences of these epitopes were identified<sup>[153]</sup> and used successfully previously in our group.<sup>[144][88][154]</sup> However, peptides cyclised with an azido di(bromomethyl)benzene (N<sub>3</sub>-DBMB) (**Figure 34**, left) handle used in the previous studies were facing the problem of poor solubility, which hampered purification, attachment to the scaffold and biological assays. Therefore, in order to increase the aqueous solubility of the cyclic peptides it was decided to use an azido triazinanetris(2-bromoethanone) (N<sub>3</sub>-TADB) (**Figure 34**, right) cyclisation handle instead of the bisbromobenzylic handle N<sub>3</sub>-DBMB. The N<sub>3</sub>-TADB handle was recently developed in our group and has shown to significantly increase the solubility of the synthesized peptides in aqueous buffers.<sup>[155]</sup> Three peptide loops cyclised with the N<sub>3</sub>-TADB hinge would mimic the discontinuous epitope in gp120, which is also built of three loops (**Figure 35**). The peptide sequences found in gp120 were mimicked by the loops with identical peptide sequences. Moreover, cysteine residues were introduced at the N- and C-termini in each of the peptides to facilitate cyclisation with N<sub>3</sub>-TADB hinge. The N<sub>3</sub>-TADB hinge is also equipped with an azide handle, which is required to click the loops onto the orthogonally protected trialkyne TAC scaffold.

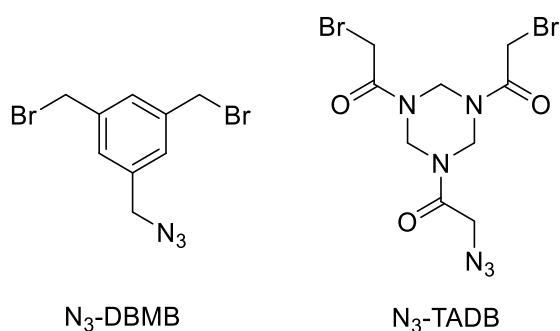


Figure 34 Structures of cyclisation hinges.  
Left: N<sub>3</sub>-DBMB cyclisation hinge. Right: N<sub>3</sub>-TADB polar cyclisation hinge.

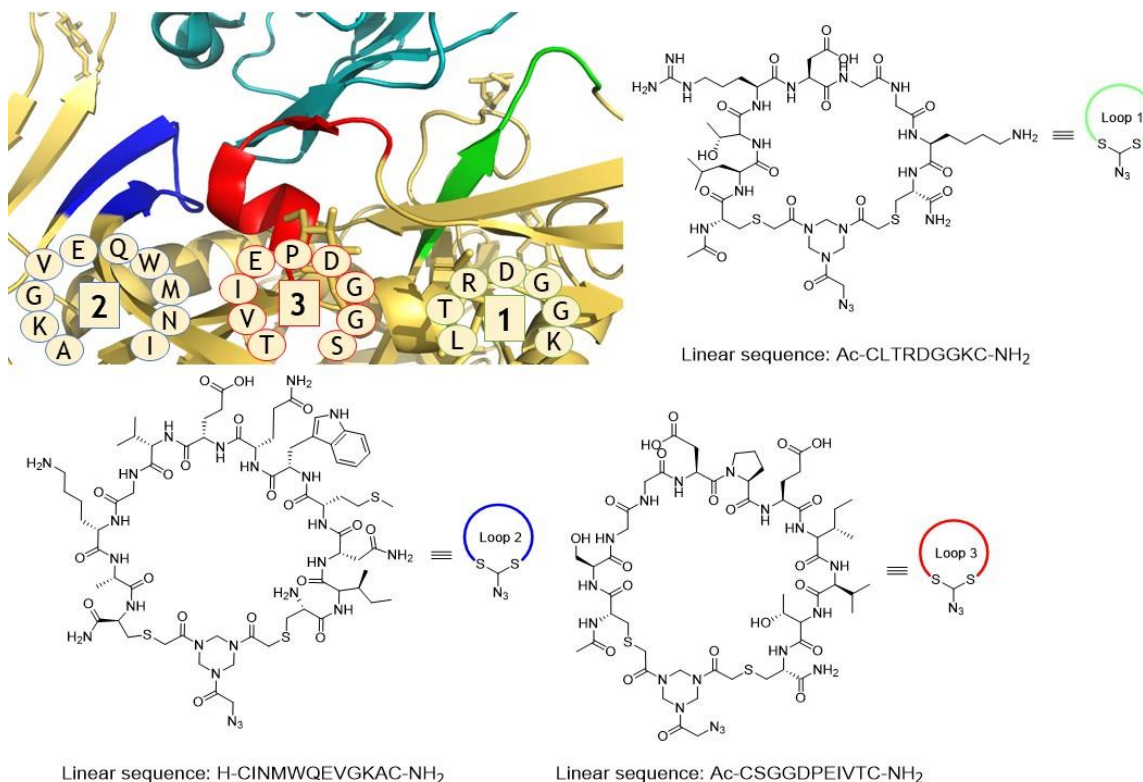
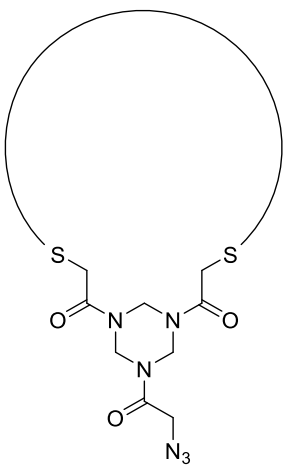


Figure 35 Synthetic peptide loops for the mimicry of gp120 CD-4 binding site. Top left: zoom on the CD4 (cyan) binding to gp120 within the conserved CD4-binding site. Loops participating in binding are shown in blue, red and green. Top right and bottom: the structures of the cyclic loops used in the mimicry of the gp120 discontinuous epitope.

The linear peptides were synthesised by means of the standard Solid Phase Peptide Synthesis (SPPS). The obtained crude peptides were cleaved off the resin and lyophilised. Since the cyclisation reaction of crude linear peptides was known to be clean and without formation of any side products, crude peptides were used directly in the reaction of cyclisation.<sup>[155]</sup> The linear peptide was dissolved together with N<sub>3</sub>-TADB linker in acetonitrile to obtain the concentration of 1 mM of the peptide. Once dissolved, 20 mM ammonium bicarbonate was added, and the cyclisation was monitored by LCMS. It was found that the cyclisation was complete after 30 min. Next, the peptide was lyophilised and purified using preparative HPLC. Three cyclic peptides corresponding to the gp120 epitope sequences were synthesized according to this protocol. The overall yields of the synthesized cyclic peptides varied from 10-29%, which corresponds to the average yield 91-95% per reaction step (Table 7).

A fraction of each of the linear peptides was separated and purified for future characterisation with <sup>1</sup>H-NMR, CD-spectroscopy and MALDI-TOF.

Table 7 The synthesized cyclic peptides based on the HIV gp120 epitope sequences and the general structure of the cyclised peptide.

Linear sequence	Overall yield (yield per step)	
Ac-CLTRDGGKC-NH <sub>2</sub> (16, Loop 1)	13% (91%)	
H-CINMWQEVGKAC-NH <sub>2</sub> (17, Loop 2)	10% (92%)	
Ac-CSGGDPEIVTC-NH <sub>2</sub> (18, Loop 3)	29% (95%)	

### 3.2.3 Attachment of the cyclic peptides to the TAC scaffold

With all the building blocks ready, it was possible to assemble the loops on the TAC scaffold.

As it was previously described by van de Langemheen et al.,<sup>[144]</sup> the cyclophane ring of the TAC scaffold is susceptible to ring-flipping, thus the “left” and “right” positions in the scaffold are interchangeable. Therefore, a construct 1-2-3 is identical to 3-2-1, where numbers indicate the peptide loop number (**Figure 35**). Since the “middle” position defines the final construct, loop 1, 2 or 3 was introduced as the first one on to the “middle” position to synthesize constructs in all possible combinations. The ligation of the peptides onto the scaffold was performed using a similar protocol to the ones published by Werkhoven et al.<sup>[88]</sup> and Longin et al.,<sup>[85]</sup> however certain modifications were introduced. It was observed, that depending on the sequence in which the loops were introduced onto the TAC scaffold, the conditions varied, especially the ratio between DMF and water as a solvent system, and reaction times (which are described in detail in Chapter 6.6). Even though structurally the constructs 1-2-3 and 3-2-1 are the same, the synthesis conditions might vary. Therefore, each of the three synthesized constructs which were used in the biological and structural studies will be discussed separately (**Figure 36**).

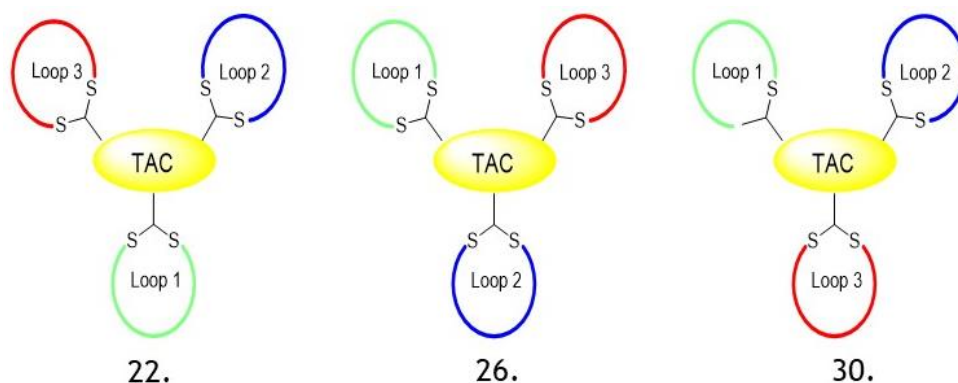


Figure 36 Schematic representation of the three final gp120 mimics used in biological and structural studies.

### 3.2.3.1 Synthesis of the construct 22 (3-1-2)

A schematic representation of the synthesis of construct **22** is shown in **Figure 37**.

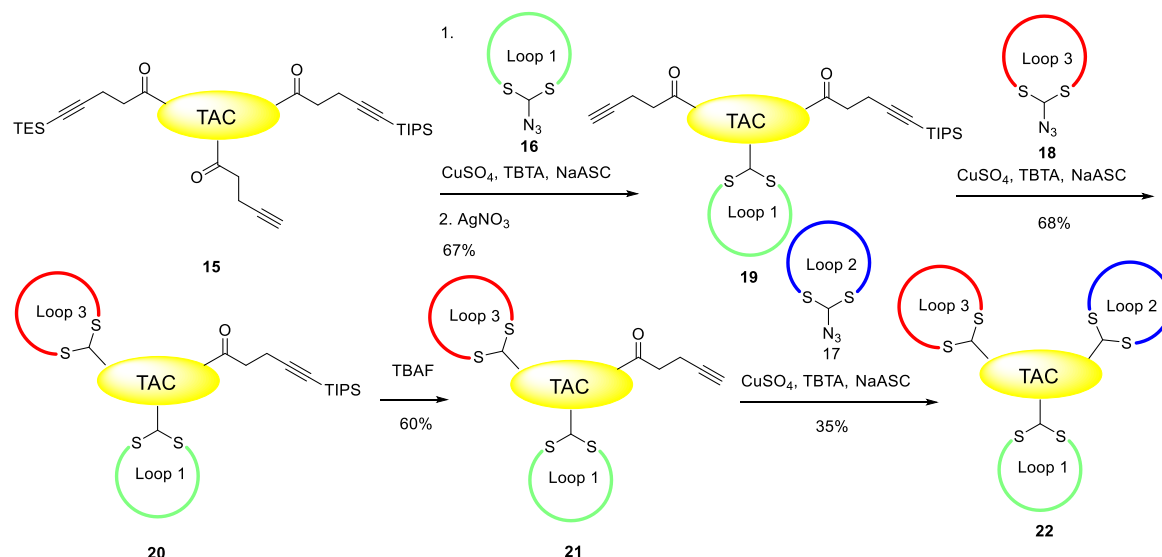


Figure 37 Schematic representation of the synthesis of construct **22**.

Loop 1 (**16**) was coupled with the free alkyne of the TAC scaffold **15** by CuAAC, which was typically completed in 2 h. This was followed by treatment with  $\text{AgNO}_3$  and under these conditions the TES group was easily removed in 45-60 min. The synthesized construct **19** was then purified by preparative HPLC and obtained in the overall yield of 67% (average yield per step: 82%). Loop 3 (**18**) was subsequently coupled to the construct **19** in 4 h and purified to afford the product **20** in 68% yield and 89% purity (the yield takes into account 89% purity of the product). Since the impurity was identified to be tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA) which did not interfere with the following TIPS-deprotection step, it was decided to proceed to the next

reaction without further purification. TIPS removal was accomplished with tetra-*n*-butylammonium fluoride (TBAF). A total of 87.5 equiv. of 1.0 M TBAF/THF had to be added to remove TIPS successfully and the reaction was completed after 29.5 h. The TIPS-removal reaction was found to be cumbersome and will be discussed in more detail in section 3.2.3.5. Once the TIPS-deprotected construct **21** was purified it was obtained in 60% yield. Lastly, loop 2 (**17**) was clicked to the scaffold bearing two cyclic peptides. This reaction was complete in 4 h and the construct **22** was obtained in a yield of 35%. Progress of the attachment of the subsequent loops to the TAC-scaffold can be followed in **Figure 38** and the final structure of the construct **22** is presented in **Figure 39**.

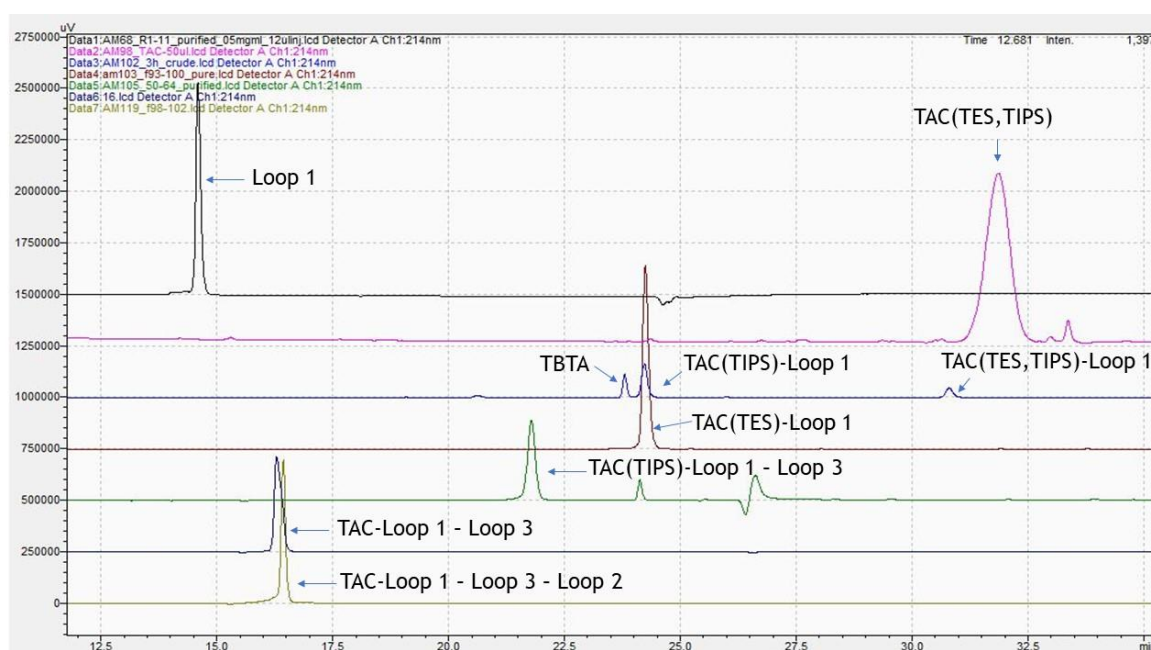


Figure 38 Progress of attachment of the peptide loops on the TAC scaffold to obtain construct **22** as monitored by analytical HPLC.

The constructs are named to describe the order in which loops were attached. Top trace to bottom trace: subsequent attachment of peptide loops to the TAC-scaffold.

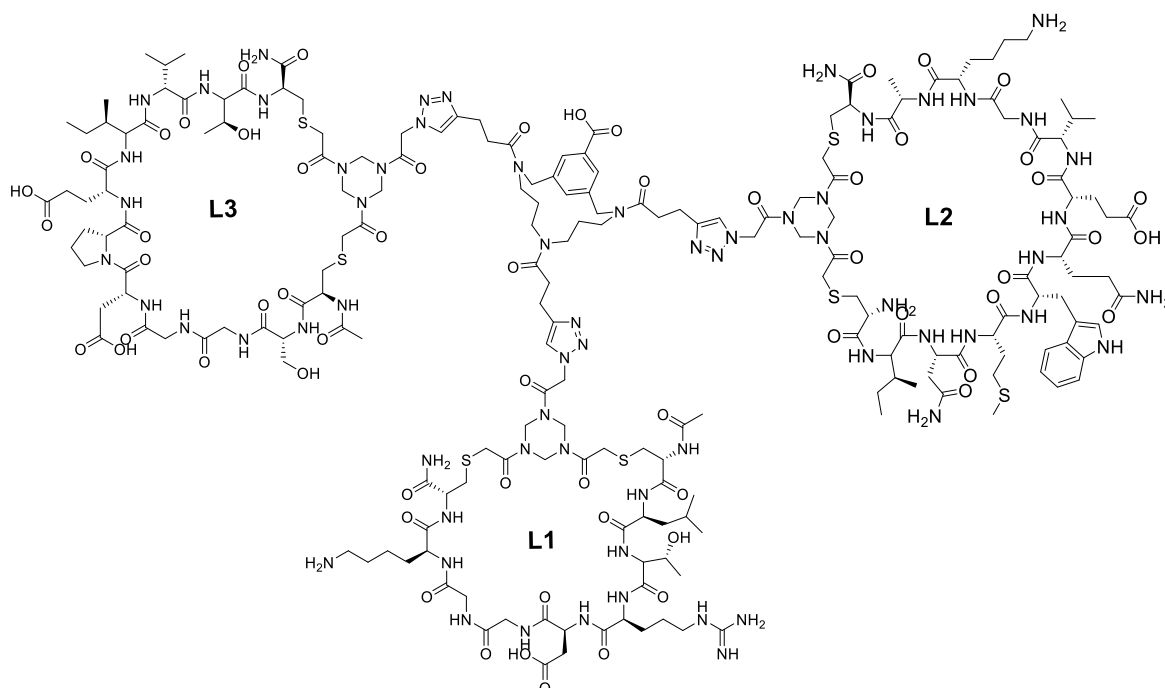


Figure 39 Structure of the final construct **22**.

### 3.2.3.2 Synthesis of the construct **30** (1-3-2):

A schematic representation of the synthesis of construct **30** is shown in Figure 40.

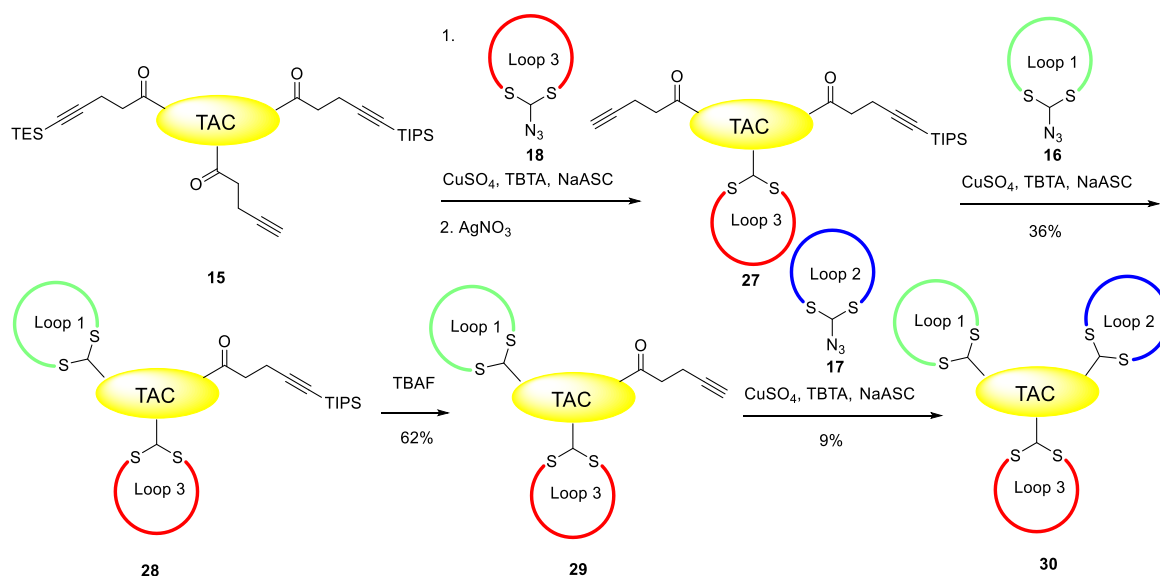


Figure 40 Schematic representation of the synthesis of construct **30**.

Synthesis of the second protein mimic was started by positioning loop 3 (**18**) in the “middle” of the TAC scaffold. The synthesis of this construct could be shortened by one purification step, which was performed for the first time after introduction of the second loop on to the TAC scaffold. In this way the synthesis time was shortened, and the number of performed purifications decreased.

Loop 3 (**18**) was coupled to TAC scaffold **15**, which was completed after 3.5 h. This was followed by TES-removal with  $\text{AgNO}_3$  in 1 h. The next step was to couple loop 1 (**16**) to the scaffold. When after 7 h the reaction was still incomplete, the reaction was left to continue overnight. After the total of 18 h, the reaction was found to be complete and could be purified to afford the product **28** in 36% yield after three steps (71% average yield per step). TIPS removal was then performed using 40 equiv. of  $\text{TBAF} \cdot 3\text{H}_2\text{O}$  in DMF and the reaction was complete in 2.5 h and after purification by preparative HPLC product **29** was obtained in 62% yield. Lastly, loop 2 (**17**) was introduced, which was accomplished after 19 h and resulted in a very low yield of 9% of the final product **30**. Progress of the attachment of the subsequent loops to the TAC-scaffold can be followed in **Figure 41** and the final structure of the construct **22** is presented in **Figure 42**.

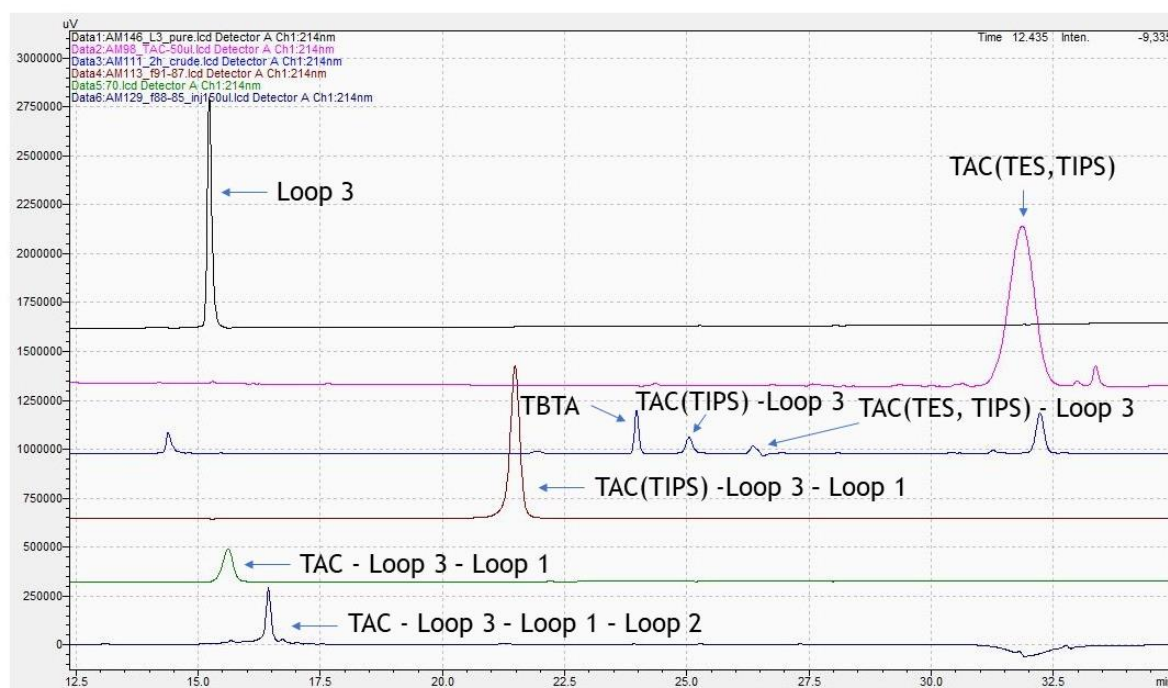


Figure 41 Progress of attachment of the peptide loops on the TAC scaffold to obtain construct **30** as monitored by analytical HPLC.

The constructs are named to describe the order in which loops were attached. Top trace to bottom trace: subsequent attachment of peptide loops to the TAC-scaffold.



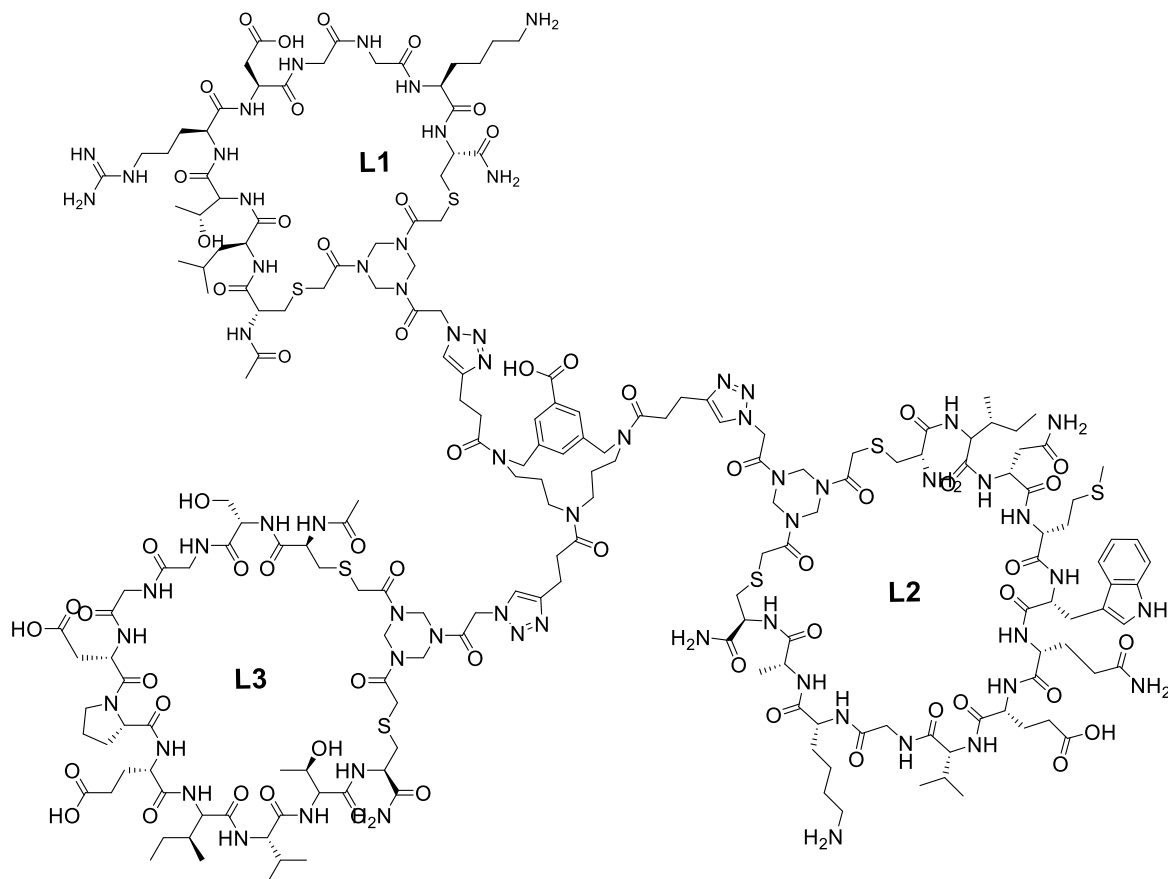


Figure 42 Structure of the final construct **30**.

### 3.2.3.3 Synthesis of the construct **26** (1-2-3)

A schematic representation of the synthesis of construct **26** is shown in **Figure 43**.

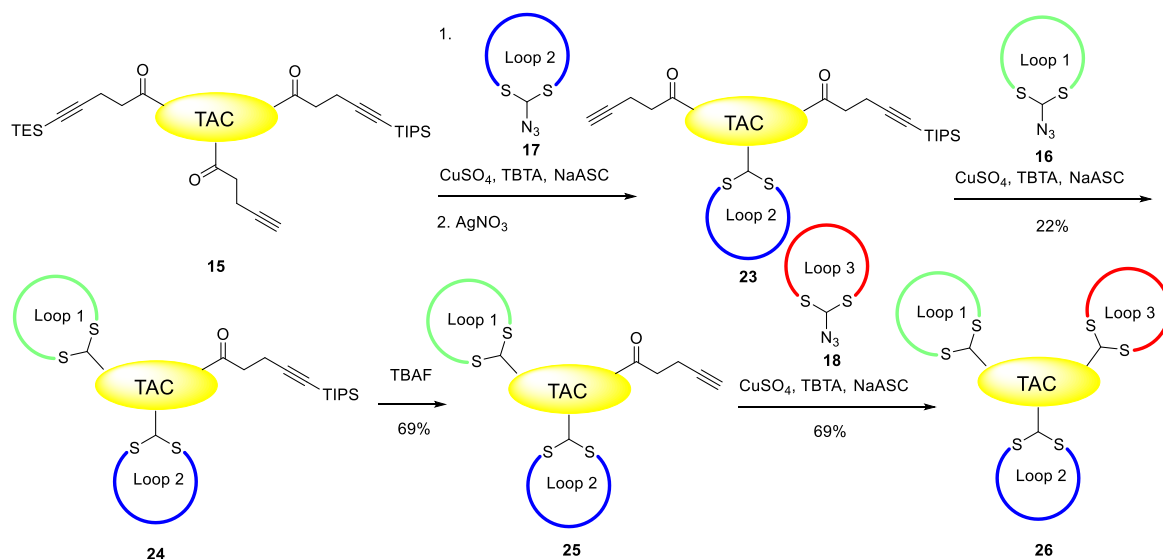


Figure 43 Schematic representation of the synthesis of construct **26**.

The synthesis of the last construct in this small library was started by coupling loop 2 (**17**) to TAC scaffold **15**, which was complete after 3 h and the subsequent



deprotection step was started and completed after 1 h. Then, loop 1 (**16**) was coupled to the construct bearing one cyclic peptide. This reaction was completed after 2 h and the product **24** was obtained in a yield of 22% over three steps (60% average yield per step). TIPS removal was performed with 40 equiv. of TBAF·3H<sub>2</sub>O in DMF and after work-up and preparative HPLC the deprotected product **25** was obtained in 69% yield. Finally, loop 3 (**18**) was coupled to the last free alkyne and this reaction was completed after 2.5 h. After purification, product **26** was obtained in a good yield of 69%. Progress of the attachment of the subsequent loops to the TAC-scaffold can be followed in **Figure 44** and the final structure of the construct **22** is presented in **Figure 45**.

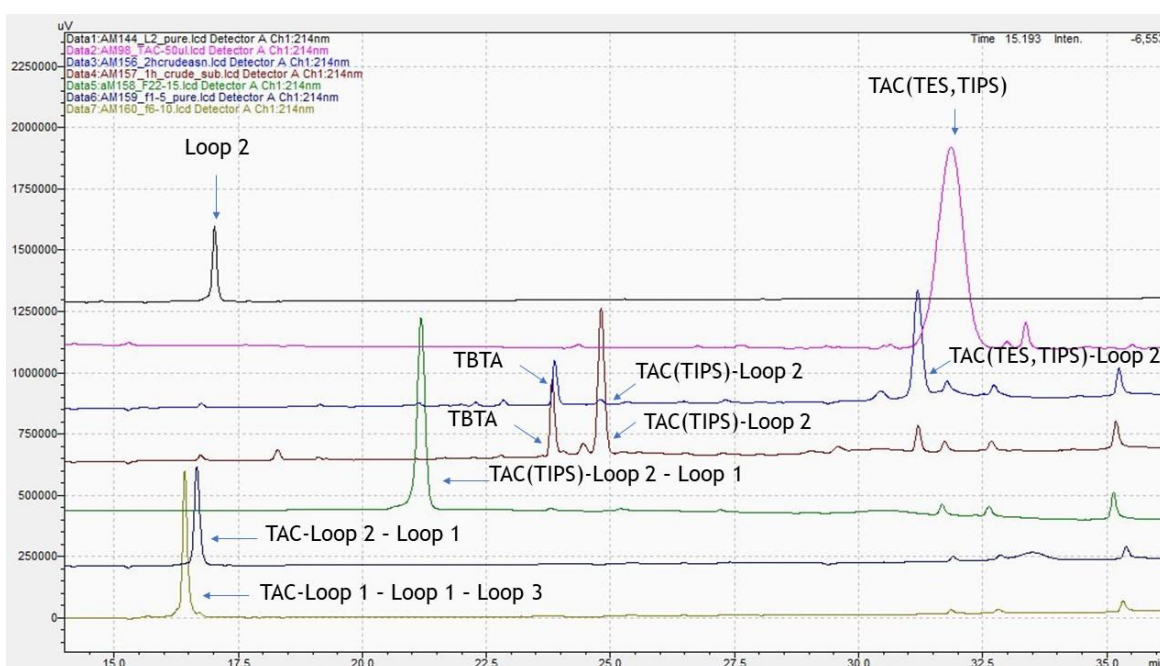


Figure 44 Progress of attachment of the peptide loops on the TAC scaffold to obtain construct **26** as monitored by analytical HPLC.

The constructs are named to describe the order in which loops were attached. Top trace to bottom trace: subsequent attachment of peptide loops to the TAC-scaffold.

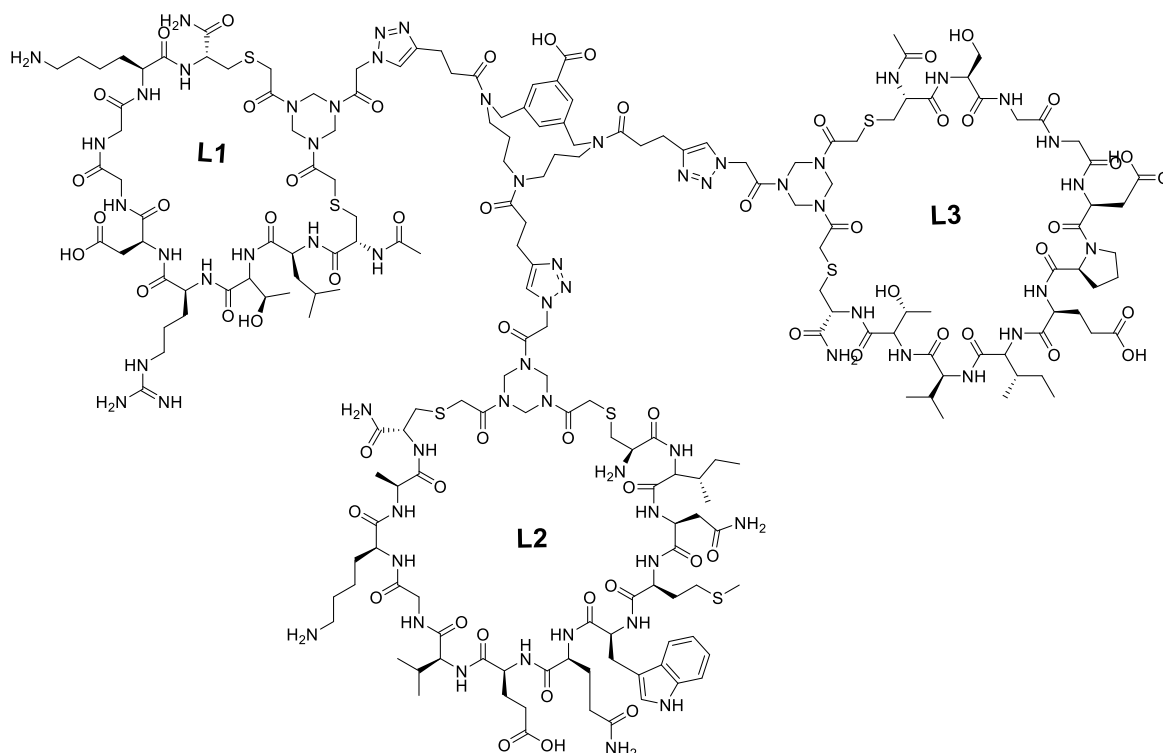


Figure 45 Structure of the final construct **26**.

Moreover, the synthesis of the construct **26** was also attempted by reversing the order in which loops 1 and 3 were introduced (**Figure 46**). Firstly, after introduction of loop 2 and TES removal as described above, loop 3 was coupled to the construct. This reaction was complete after 10 h and proved to be difficult to monitor, since the product peak overlaps with the TBTA peak on LCMS. Also, because the product and TBTA run closely together on the chromatography column, the product could not be purified and was obtained as a mixture with TBTA with the yield of 11% after three steps. TIPS-removal with 40 equiv. of TBAF·3H<sub>2</sub>O was complete after 4.5 h with 54% yield. Finally, loop 1 was introduced, which was complete after 3 h and after the purification the final product **26** was obtained in 23% yield. However, only 0.6 mg of this product was obtained, which was mostly due to the loss of the product after introduction of the second loop. Due to the small amount of the obtained product, it was not used for structural or biological studies. The progress of this reaction was followed by analytical HPLC (see Appendix **Figure 146** - **Figure 157**).

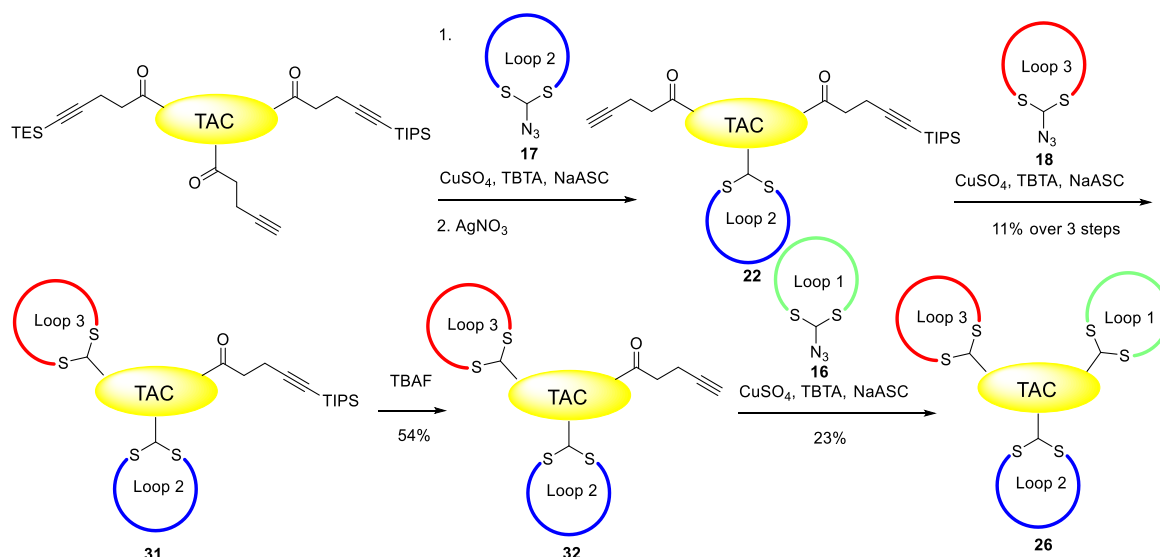


Figure 46 Schematic representation of the synthesis of construct **26** with different order of loops attachment.

### 3.2.3.4 Overview of loops attachment to TAC-scaffold

The attempt to synthesize compound **26** by introducing the loops in different order clearly demonstrated, that the order in which the loops were introduced onto the scaffold was of enormous importance. Another explanation for this difference was the slightly more hydrophobic character of loop 3. When comparing the retention times on analytical HPLC of loops 1 and 3, which were 14.593 min and 15.227 min (gradient of A into B of 0-100% over 30 min), respectively, it was observed that loop 3 was slightly more hydrophobic. Loop 2 was known to be less soluble and more hydrophobic when compared to loops 1 and 3, with the retention time on analytical HPLC of 17.023 min. Moreover, the loops are of different sizes, since loop 1 consists of 9 amino acids, loop 2 of 12 amino acids and loop 3 of 11 amino acids. The hydrophobic character and size of the loops are schematically shown in Figure 47.

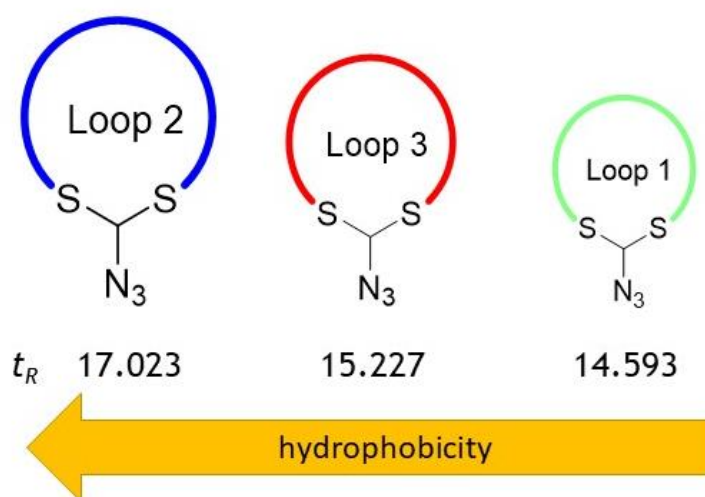


Figure 47 Comparison of size and hydrophobic character of independent loops with the retention times obtained from analytical HPLC.

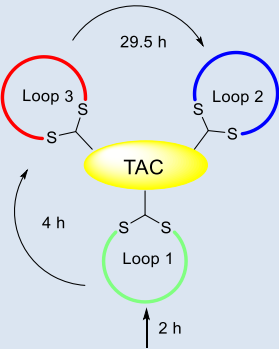
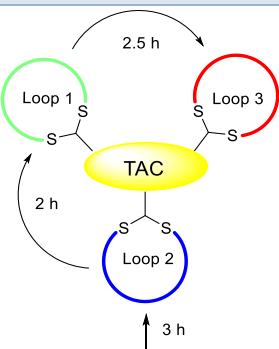
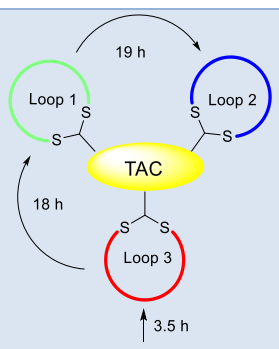
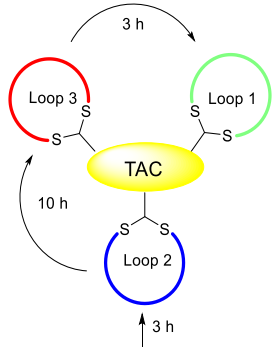
It is believed, that the final yield of the gp120 mimics is a result of hydrophilic/hydrophobic character of the loops, the amino acid composition in the primary structure and the order in which they are introduced (**Table 8**).

Introduction of loop 1, which is the smallest of the loops (shortest peptide sequence) and the most hydrophilic, which is followed by loop 3 (medium sized, medium hydrophilicity) and then loop 2 (the biggest and most hydrophobic) resulted in a good yield of the final product **22**. This might indicate, that loops 1 and 3 interacted with each other and “left enough space” for an efficient, albeit time-consuming, introduction of loop 2 (**Table 8**, entry 1). If loop 2 was introduced as the first one, it had the most space to be coupled efficiently. When followed by loop 1, which is the smallest of the loops, again enough space is left for loop 3. This combination resulted in the highest overall yield of the compound **26** (**Table 8**, entry 2). However, when the synthesis is started from loop 3 and followed by loop 1 (reverse order when compared to entry 1), it seemed that the loops might be interacting with each other differently, since the final overall yield is significantly lower and coupling times longer (**Table 8**, entry 3). When the synthesis was started from loop 2 and followed by loop 3, which are the two biggest loops, it seemed that the free alkyne in the “right” position was hindered significantly, and loop 1 could not be coupled efficiently, resulting in the lowest overall yield of 1.3% (**Table 8**, entry 4). However, the yields of each loop introduction could not be compared, because in two out of three constructs two loops were introduced in a one-pot reaction. It was also noted, that depending on

the position on the scaffold in which the loop is being attached, the coupling times vary. For example, introduction of loop 3 followed by loop 1 was accomplished in 18 or 3 h (**Table 8**, entry 3 and 4, respectively), depending on the position of the loops on the scaffold. However, introduction of loop 1 followed by loop 3 was complete in 4 or 2.5 h (**Table 8**, entry 1 and 2, respectively). All these observations indicated, that several variables, including hydrophilic/hydrophobic character of the loops, the amino acid composition in the primary structure, the order in which they were introduced have influenced the final yield of the obtained product.

The secondary structure of the peptides and their interactions with each other when assembled on the scaffold were studied in more detail by <sup>1</sup>H-NMR and CD-spectroscopy and are described in section 3.3.

Table 8 Synthesized constructs with attached loops in different positions, yields and amounts in which they were obtained. The arrow in the schematic structure indicates the order in which the loops were introduced, with the arrow starting as the first one.

Entry (compound)	Amount obtained	Overall yield	Schematic structure
1. (22)	3.2 mg (0.6 $\mu$ mol)	9.4%	
2. (26)	3.4 mg (0.7 $\mu$ mol)	10.6%	
3. (30)	1.2 mg (0.2 $\mu$ mol)	2.3%	
4.	0.6 mg (0.1 $\mu$ mol)	1.3%	

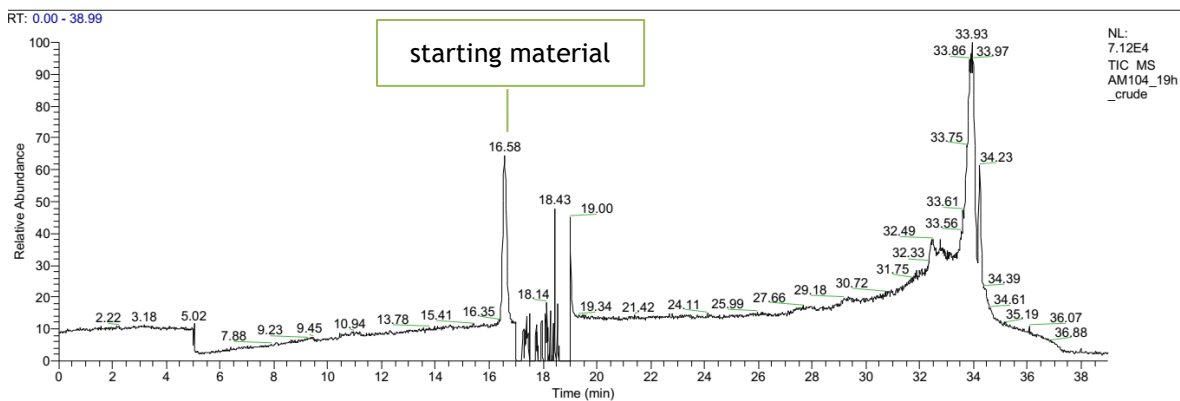
### 3.2.3.5 TIPS-removal

The TIPS-protecting group has proven to be very stable and even difficult to remove. Removal conditions were screened to find the best method and the reactions were monitored by LCMS or analytical HPLC. Selected examples will be shown below. A very fine balance between the number of TBAF equiv. and reaction times had to be found. Extended reaction times and too small amount of equiv. of TBAF led to the decomposition of both product and starting material.

In one example, compound **28** served as a starting material. After 19 h from addition of 30 equiv. 1 M TBAF/THF, only starting material was observed by LCMS (**Figure 48**). After addition of another 18 equiv. and a further 4 h, there was still no conversion (**Figure 49**), therefore another 18 equiv. of TBAF were added. After a total of 26 h the formation of the product was observed, however, the majority still being starting material (**Figure 50**). 17.5 h later the ratio between the product and the starting material started shifting even further towards formation of the product. To speed up the reaction, another 18 equiv. of TBAF were added, however after next 6.5 h no further conversion was observed. Addition of another 18 equiv. of TBAF and leaving the reaction mixture for 16.5 h more resulted in the complete decomposition of both starting material and the product (**Figure 52**).

C:\Users\...\TBAFs\AM104\_19h\_crude

17/04/2018 08:57:56



AM104\_19h\_crude #982 RT: 16.58 AV: 1 NL: 2.72E3  
T: ITMS + p ESI Full ms [150.00-2000.00]

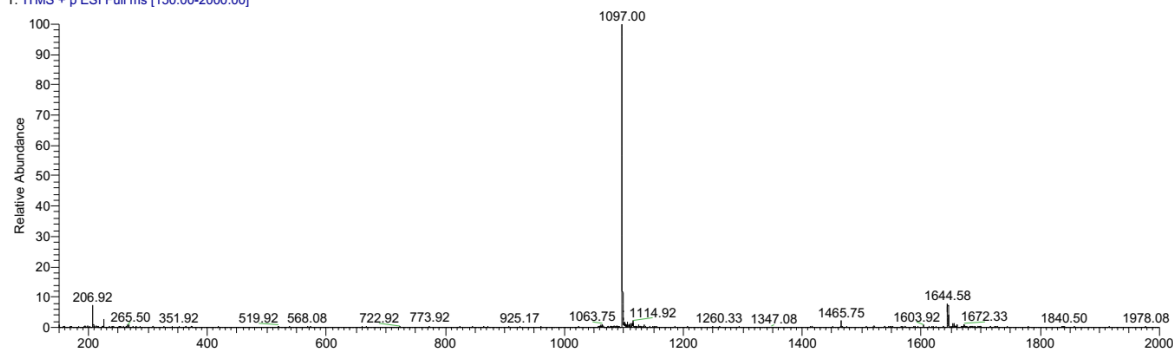


Figure 48 Progress of TIPS removal after 19 h.

C:\Users\...\TBAFs\AM104\_23h\_crude

17/04/2018 14:14:17

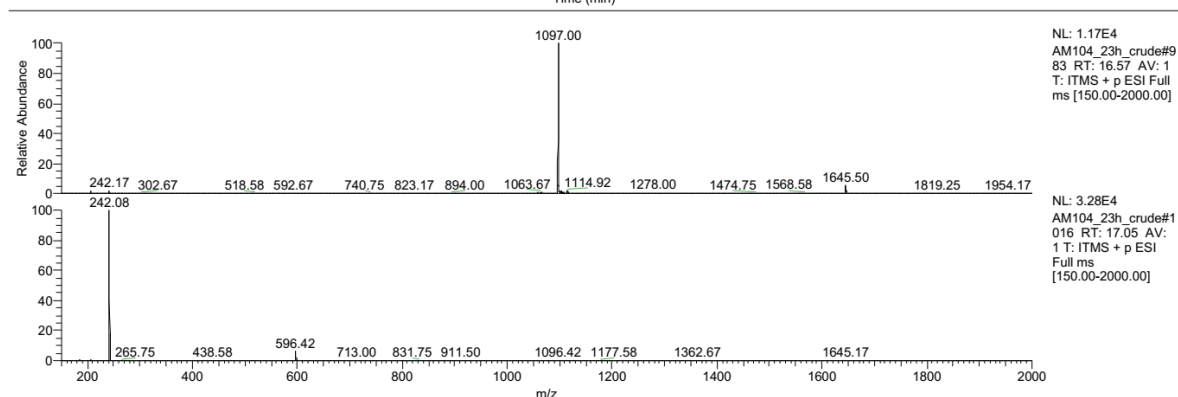
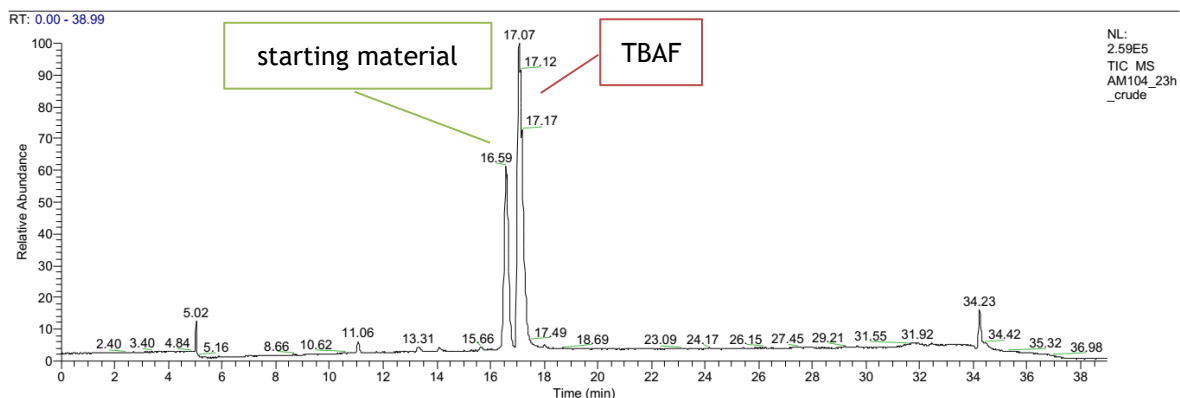


Figure 49 Progress of TIPS removal after 23 h.



C:\Users\...\TBAFs\AM104\_26h\_crude

17/04/2018 16:22:27

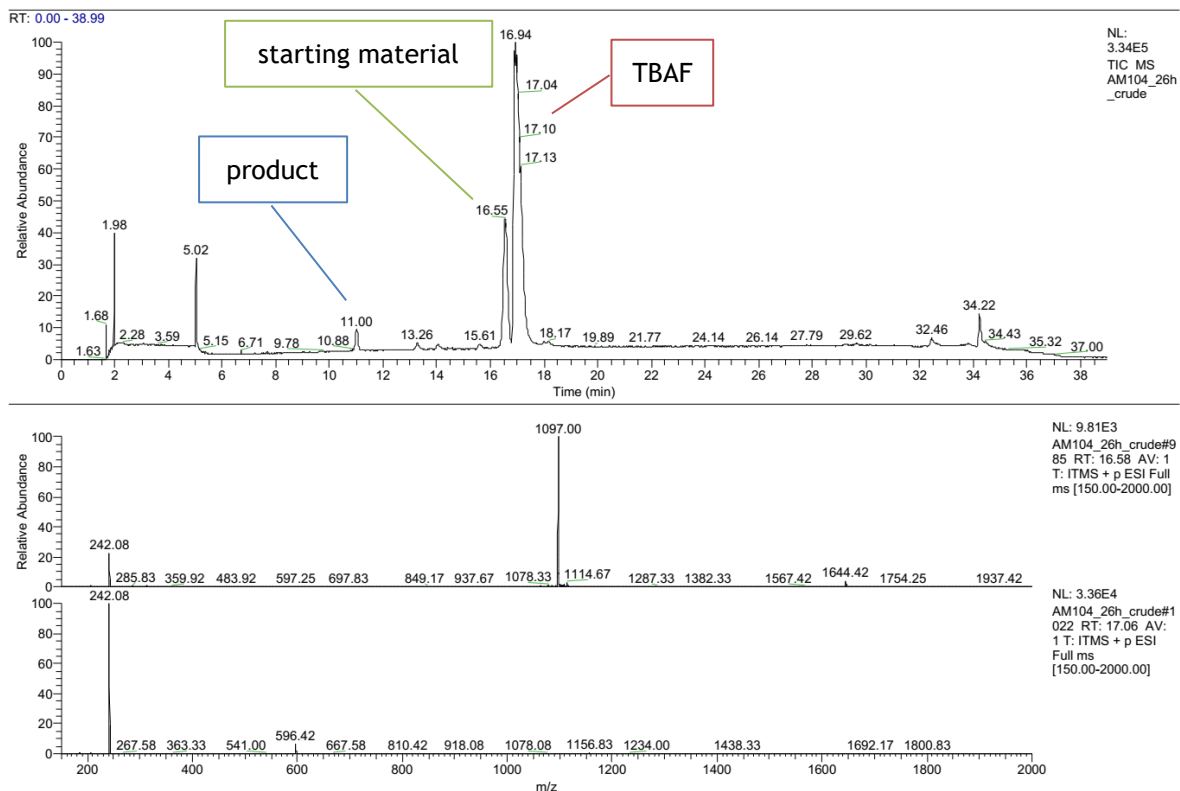


Figure 50 Progress of TIPS removal after 26 h.

C:\Users\...\TBAFs\AM104\_43.5h\_crude

18/04/2018 09:31:48

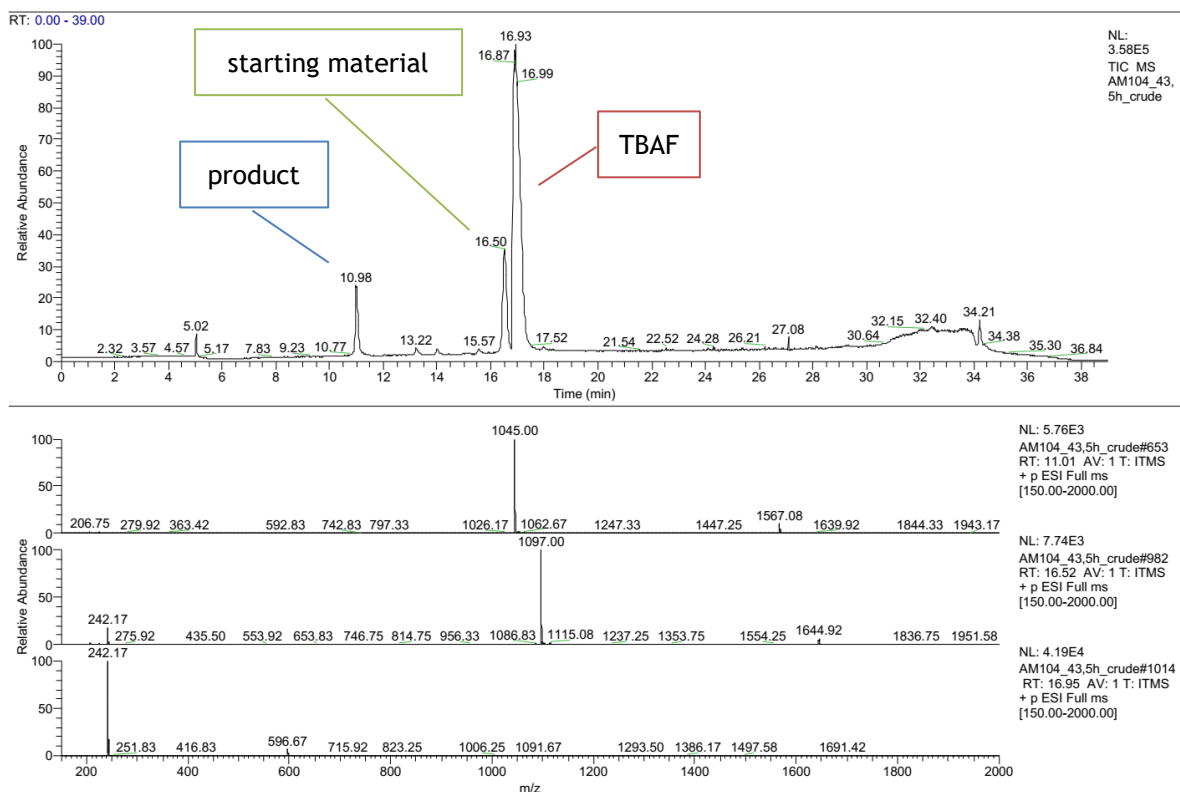


Figure 51 Progress of TIPS removal after 43.5 h.

C:\Users\...\TBAFs\AM104\_66,5h\_crude

19/04/2018 08:07:13

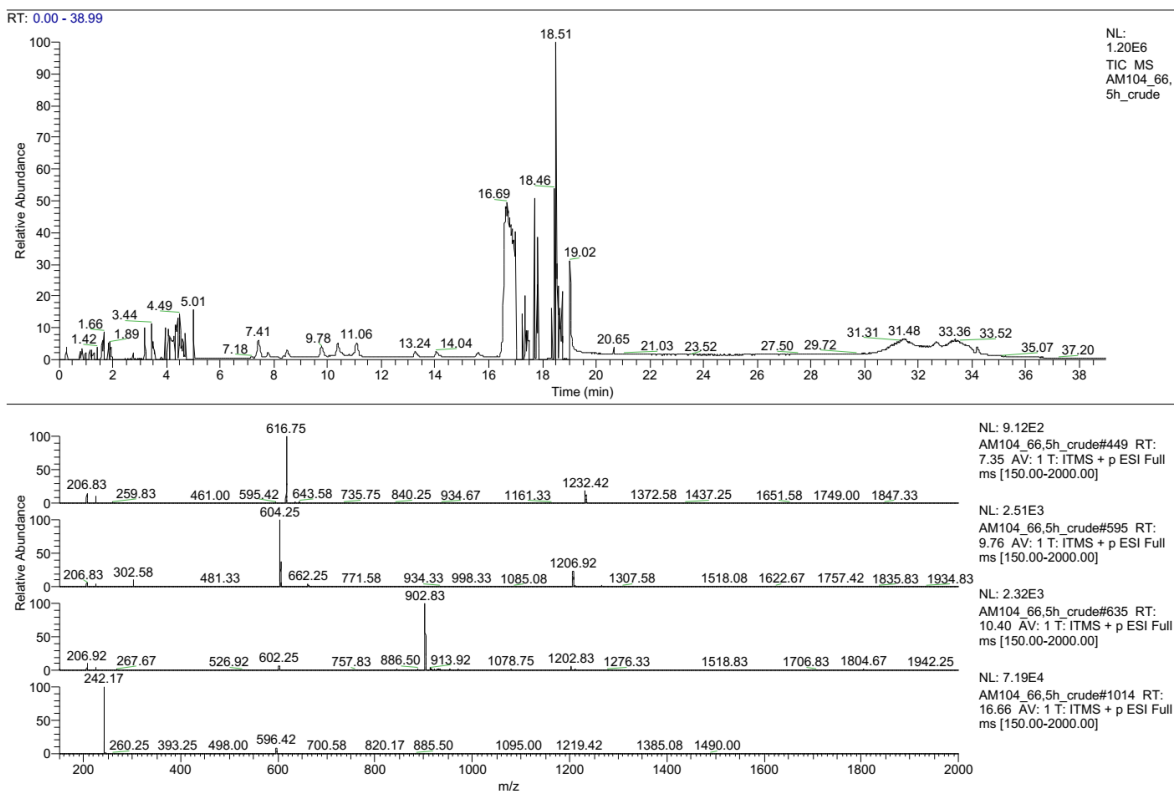


Figure 52 Product decomposition after the total of 66.5 h.

In contrast to the above example, it was found that the reaction led to the formation of the desired product in 2-4.5 h depending on the composition of loops on the scaffold, if between 40-50 equiv. of TBAF·3H<sub>2</sub>O was added all at once instead of adding it in portions (**Figure 53**). TBAF·3H<sub>2</sub>O was preferred to 1 M TBAF/THF, because it has less basic character and a smaller chance of causing decomposition and racemisation in peptides.

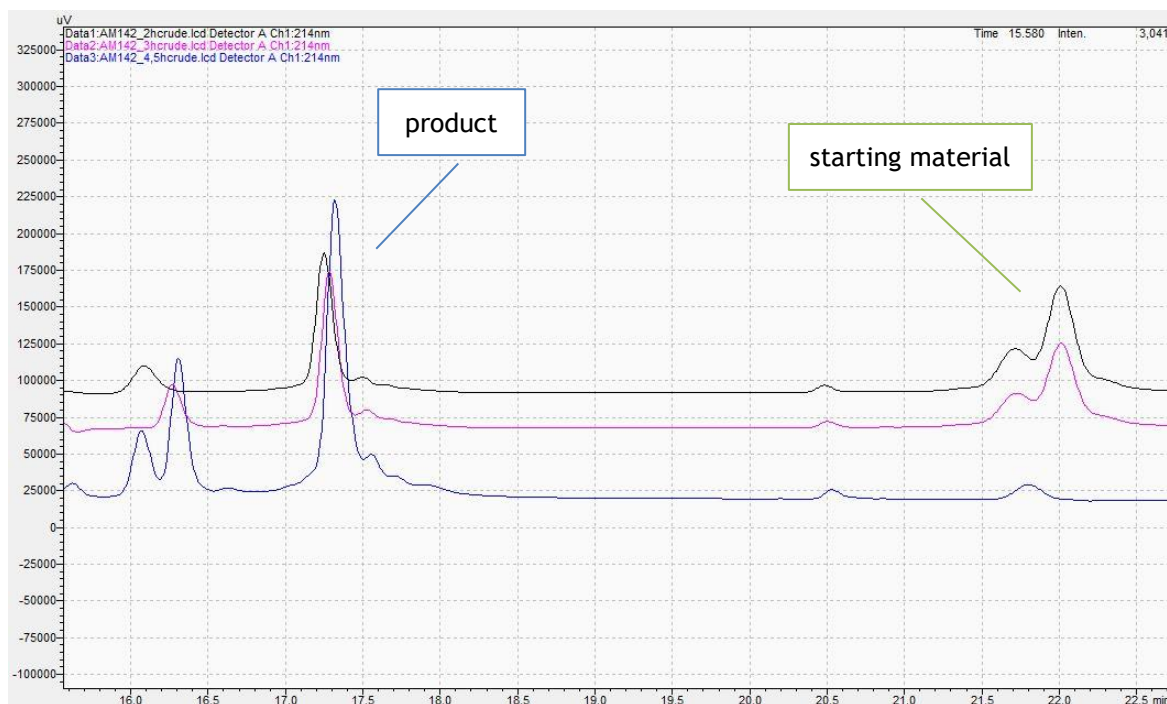


Figure 53 Zooming in on progress of the TIPS deprotection reaction with TBAF·3H<sub>2</sub>O over time monitored by analytical HPLC.

Black line: reaction after 2 h, pink line: reaction after 3 h, blue line: complete conversion after 4.5 h.

### 3.2.3.6 One-pot assembly of the peptide loops on the TAC scaffold

A successful attachment of two peptide loops and removal of TES-protecting group in one-pot allowed to shorten synthesis time and decrease the number of purification steps. This inspired us to investigate whether it was possible to assemble all loops in a one-pot reaction. A single attempt towards one-pot attachment of the loops and removal of the protecting-groups was performed. It was not possible to repeat this reaction due to the limited availability of the starting materials.

Attachment of the first peptide loop (loop 2), TES-removal and attachment of the second peptide loop (loop 1) in one-pot was followed by analytical HPLC and was performed successfully (for the reaction scheme, see Figure 43. For analytical HPLC chromatograms monitoring the reaction see Appendix Figure 158 - Figure 160). The bottleneck reaction proved to be TIPS-removal. The progress of this reaction can be followed in Figure 54. After 2 h from addition of 34.2 equiv. of TBAF·3H<sub>2</sub>O the ratio between the product and the starting material was 20:80%. After addition of 18 equiv. of TBAF·3H<sub>2</sub>O and further 1 h the ratio shifted to 34:66%. After addition of further 21 equiv. and another 1 h, formation of a side

product was observed as well as decomposition of the starting material. The ratio between the side product, product and (decomposed) starting material was: 34%:25%:41%. After leaving the reaction overnight, neither product nor starting material could be detected.

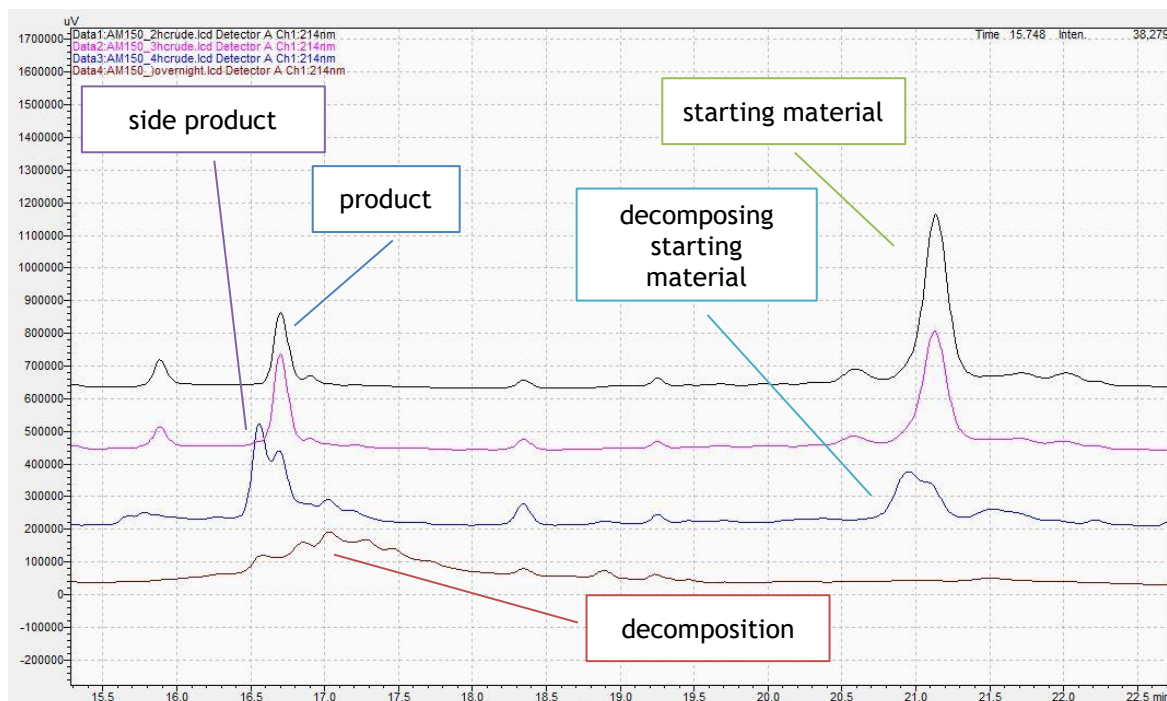


Figure 54 Progress of the TIPS-removal in one-pot assembly of the discontinuous mimics monitored by analytical HPLC.  
Black line: reaction after 2 h, pink line: -reaction after 3 h, blue line: reaction after 4 h, brown line: reaction after overnight.

LCMS analysis (Figure 55) of the reaction mixture after 4 h had shown that the mass of the side product was lower than the mass of the desired product by 174 mass units. The same difference was observed between the starting material and its decomposed variant. It shows that both constructs undergo a similar decomposition pathway caused by presence of TBAF which might be caused by  $\beta$ -elimination. The base could abstract the  $^{\alpha}\text{H}$  of the Cys, which is coupled to the cyclisation linker. If the sulphur is oxidised, which might happen in presence of remains of cooper, as it is in the click reactions,  $\beta$ -elimination would be even easier. A possible remedy to this undesired reaction would be to exchange cysteine to homocysteine, which would be less susceptible to the  $\beta$ -elimination.

This example showed, that if the reaction conditions were carefully optimised, it would be possible to remove TIPS-protecting group in presence of the click reagents. This would make the one-pot synthesis feasible. However, as it was

shown in the previous sections, that the amount of TBAF is of critical importance, as well as the time for how long the reaction is performed.

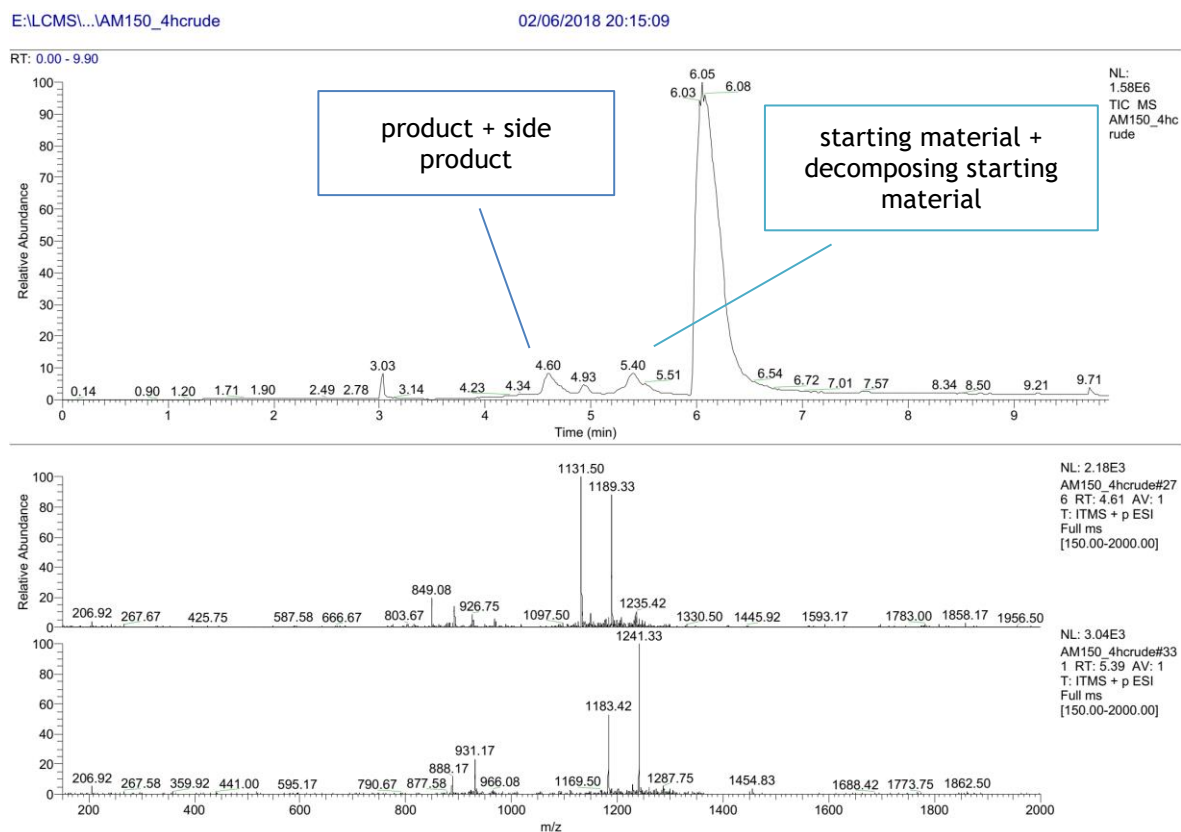


Figure 55 LCMS analysis of the TIPS removal reaction in one pot after 4 h.

### 3.3 CD and NMR-spectroscopy towards structural characterisation of the synthesized gp120 mimics and independent peptide loops

Linear peptides corresponding to the primary sequence of gp120 discontinuous epitopes, their cyclised versions as well as the total constructs with loops assembled on the TAC scaffold were analysed by NMR and CD-spectroscopy. These experiments allowed to gain insight into the secondary structure of the synthesized molecules, the amount of the structural organisation, and if and how the loops were interacting with each other once assembled on the TAC scaffold.

### 3.3.1 <sup>1</sup>H-NMR-spectroscopy analysis of synthesized linear and cyclic peptides and assembled constructs

The NMR experiments were performed and analysed with the assistance of Dr. Brian Smith (Institute of Molecular, Cell and Systems Biology University of Glasgow).

The samples were prepared as 100 µM solutions of the compound in 20 mM phosphate buffer, pH 7.5, supplemented with D<sub>2</sub>O to a final concentration of 5%, total volume 600 µL. However, both accuracy of weighing out these molecules (less than 0.3 mg per sample, which was further diluted to obtain the desired concentration) as well as problems with their precipitation in the aqueous buffer (especially linear peptide 2) might have caused that the actual concentration was lower than expected. Moreover, linear peptides were prepared both with addition and without TCEP as reducing agent to compare, whether the linear peptides are present in reduced or oxidised form in a solution. It was found however, that addition of TCEP did not influence significantly the obtained spectra, which suggests that the linear peptides are in the reduced state. Selected spectra are shown and analysed below. Since spectra of the linear peptides with and without TCEP are almost identical, only the spectra without TCEP are shown below. Spectra of the TCEP treated linear peptides are shown in the Appendix (**Figure 161 - Figure 163**). The water signal was suppressed in all the spectra.

The NMR spectra of peptides can be analysed similarly to proteins, as it was described in the section 2.1.4.

The same samples were used for the CD-spectroscopy experiments.

#### 3.3.1.1 Linear peptides

<sup>1</sup>H-NMR analysis of the linear peptide 1 suggested, that it was present in a random coil configuration (**Figure 56**). Very weak signals coming from the amide backbone NH protons mean that these amide protons are rapidly exchanging with the solvent and are not protected from the environment. This observation suggests that linear peptide 1 is mostly disordered. The spectrum of an unfolded peptide corresponds to the spectrum of a sum of random coil spectra of the amino acid residues of which the peptide is composed.<sup>[156]</sup> The sharp and not very well dispersed peaks

in the aliphatic region allow to find some of the characteristic shifts of the amino acids (all values given in ppm): Leu:  $\delta$  CH<sub>3</sub>: 0.87, 0.94; H <sup>$\beta$</sup> : 1.62, 1.64; Thr  $\gamma$ CH<sub>3</sub>: 1.21; Arg  $\delta$ CH<sub>2</sub>: 3.20, 3.19; Lys  $\epsilon$ CH<sub>2</sub>: 3.00, 2.98. However, <sup>1</sup>H-NMR experiment did not allow to assign all the protons found in the spectrum.

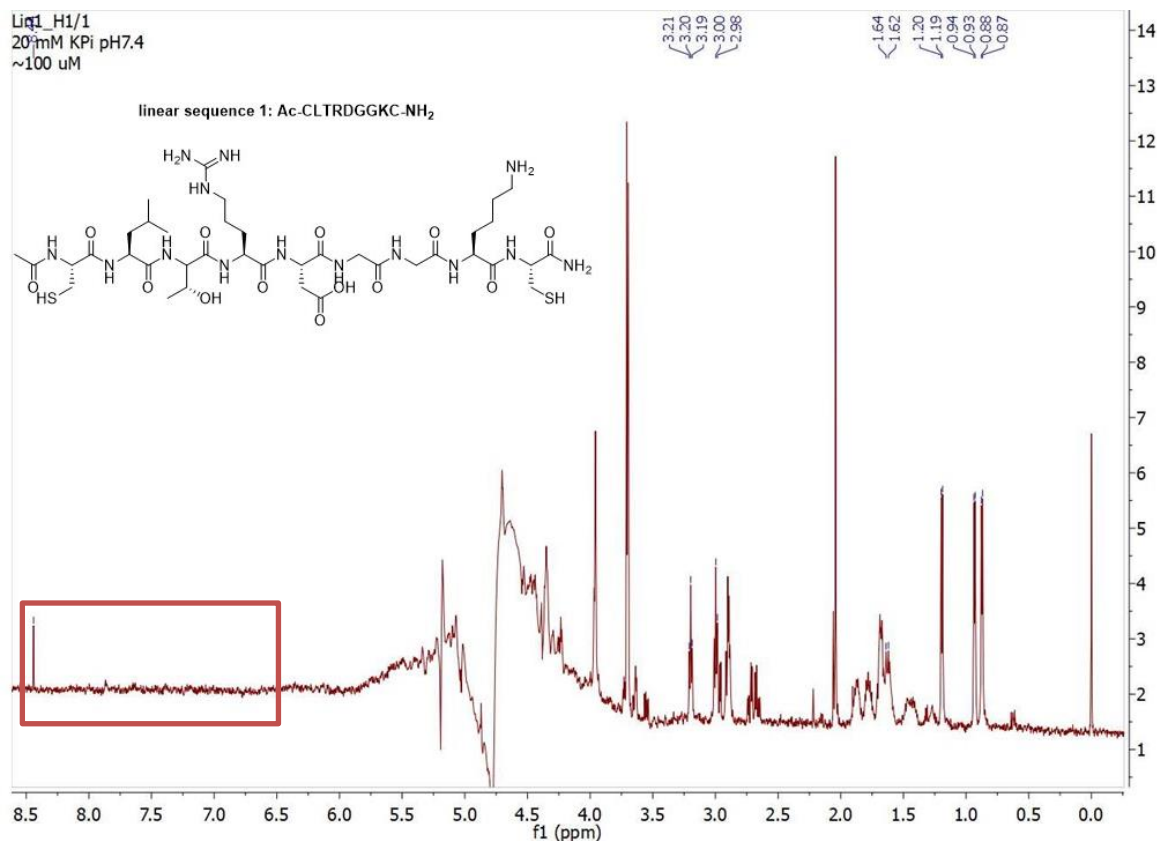


Figure 56 <sup>1</sup>H NMR analysis of linear peptide 1. The red box indicates where the amide backbone proton signals should be present if the linear peptide 1 would have had ordered secondary structure.

The linear peptide 2 heavily precipitated out in the buffer, therefore the signal:noise ratio in the NMR spectrum is low (**Figure 57**). The most characteristic is the peak present at 10.16 ppm, which belongs to the proton of the amine in the indol ring of the tryptophan residue. Moreover, the aromatic protons of the tryptophan residue are visible in the 7.00-8.00 ppm region. Overall, more dispersed peaks in the aliphatic region, when compared to the linear peptide 1, suggest that linear peptide 2 is more structured. The signals coming from the NH protons of the amide backbone indicated that these are more protected from the solvent environment, which could be caused by the fact that this peptide is more constrained already in its linear form.



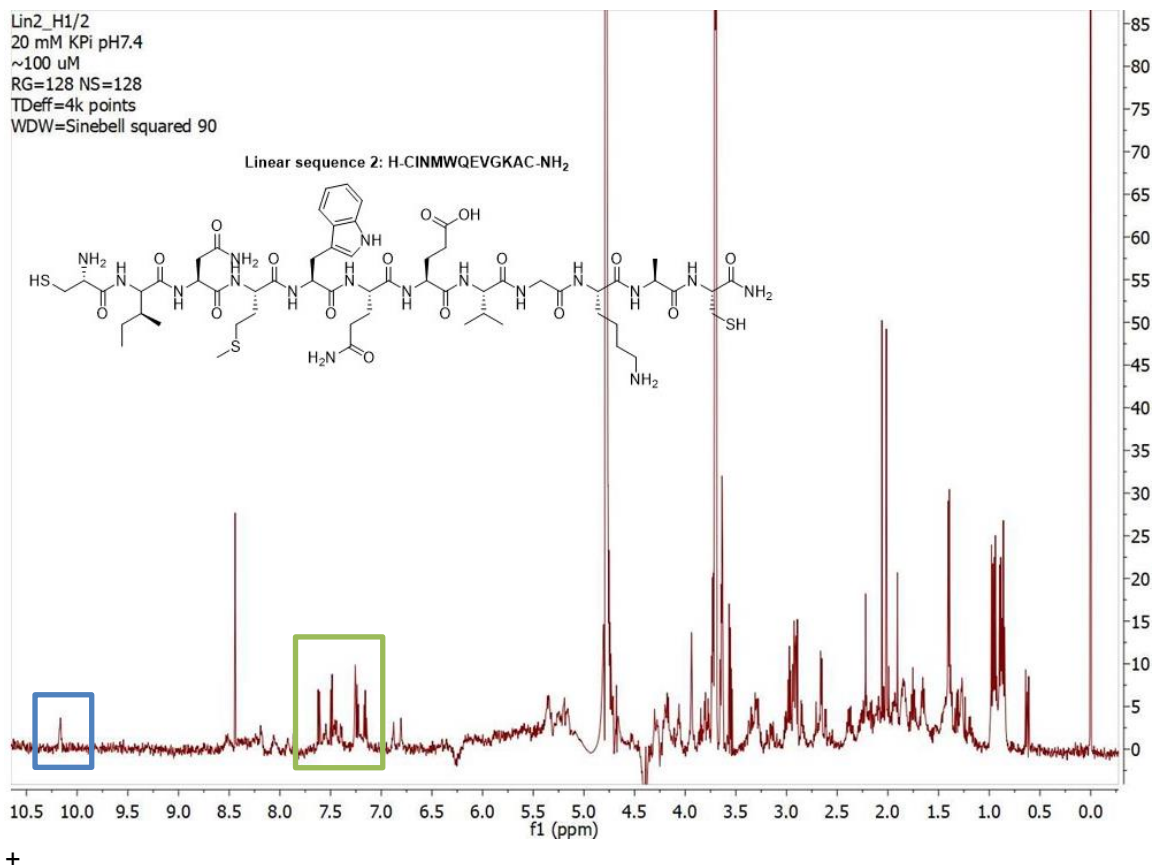


Figure 57 <sup>1</sup>H NMR analysis of linear peptide 2.

Green box indicates aromatic protons belonging to the indol ring of Trp. The blue box indicates the NH proton from the indol ring.

In the <sup>1</sup>H-NMR spectrum of the linear peptide 3 (**Figure 58**) the signals coming from NH protons of amide bonds are very strong (8.0-8.5 ppm), which indicates that these protons are protected from the surrounding solvent and the amide backbone is more constrained. Linear peptide 3 does not have any aromatic amino acids in its primary sequence nor primary NH<sub>2</sub> groups as in Asp, Glu or Lys, therefore the signals present in the 7.0-7.5 ppm region originates from the NH<sub>2</sub> group of the C-terminus. Linear peptide 3 seems to have the most constrained structure out of the three linear peptides.



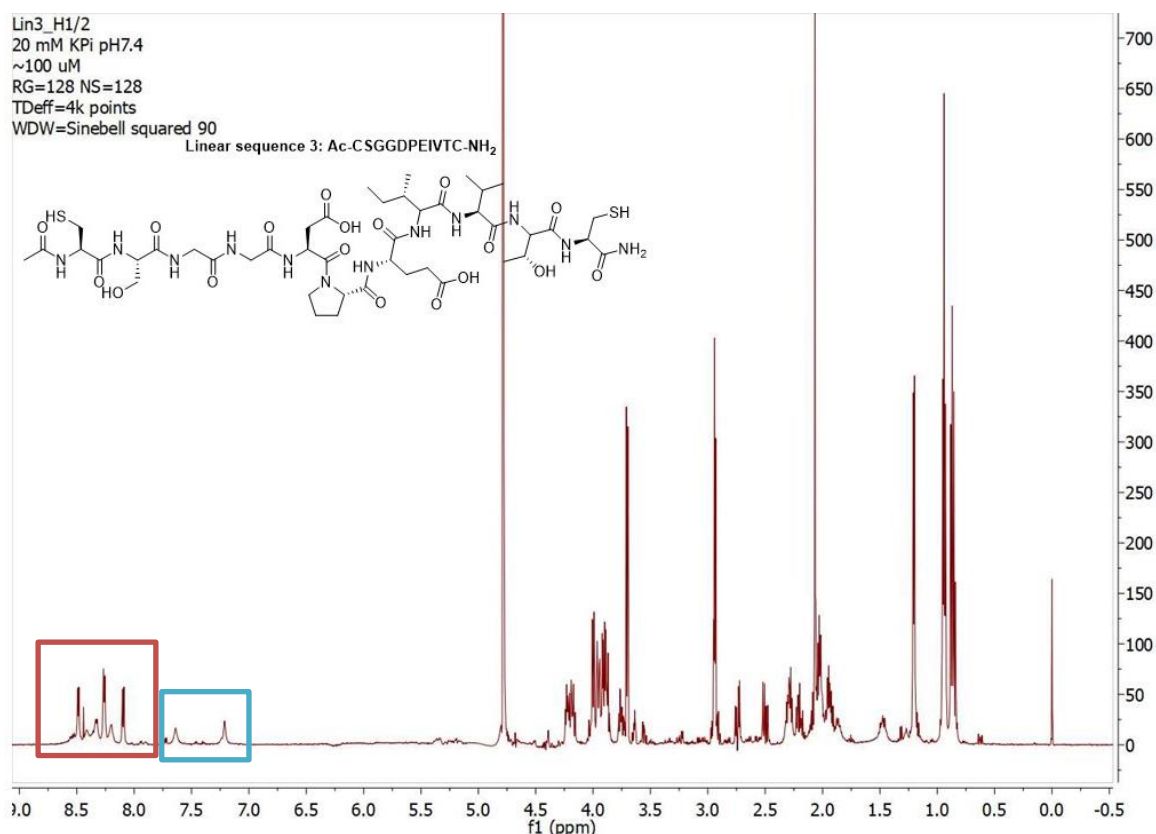


Figure 58 <sup>1</sup>H NMR analysis of linear peptide 3.

The red box indicates the proton signals coming from the amide backbone of the peptide suggesting that it obtains a certain secondary structure. The cyan box indicates protons signals coming from the C-terminal NHs.

### 3.3.1.2 Cyclic peptides

In the <sup>1</sup>H-NMR spectrum of loop 1 small changes in the amide region were observed (**Figure 59**). The clustered amide NH protons signals in the 8.0-8.5 ppm region indicate a random coil structure in this peptide. Moreover, the proton signals of Arg and Lys NH protons can be observed. The aliphatic region representing the side chains of the peptide is almost identical to the linear peptide.

If the two spectra are stacked over each other, it can be observed that there are differences in the amide backbone region (**Figure 60**). This suggests that even though both peptides are disordered, they adopt different conformations.

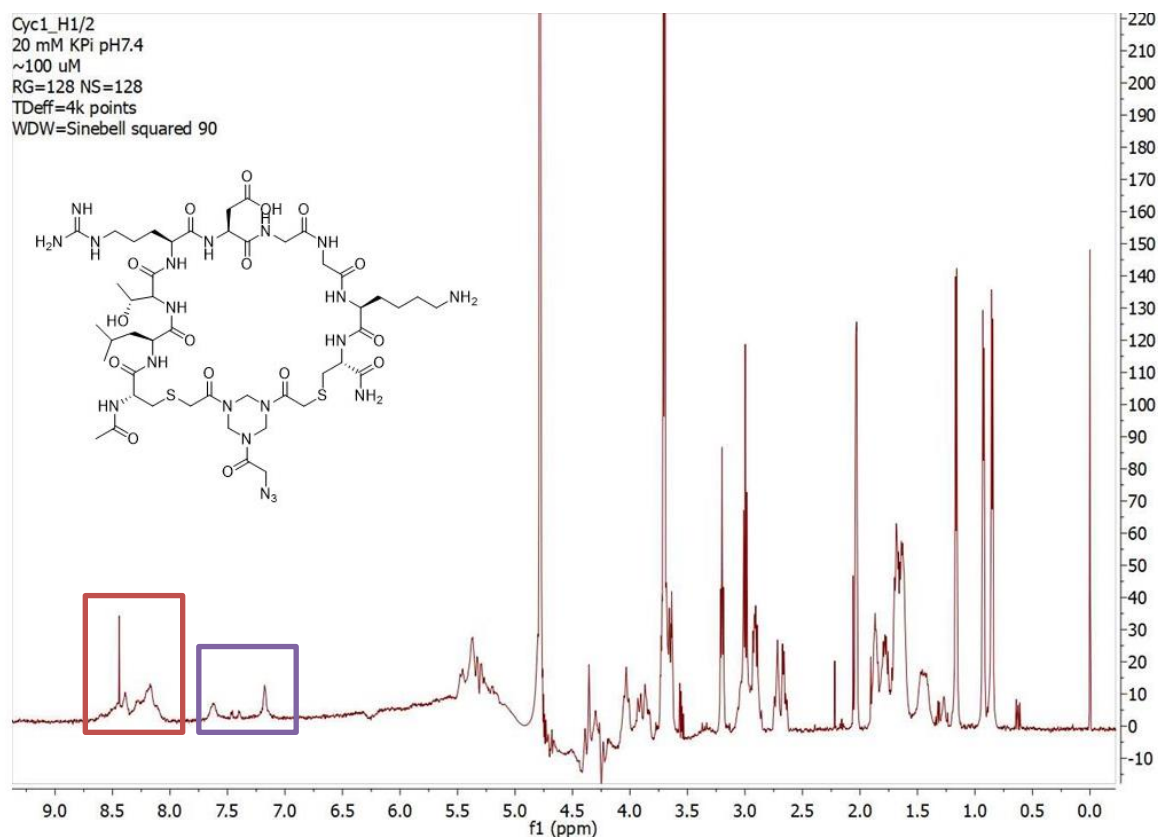


Figure 59  $^1\text{H}$  NMR analysis of Loop 1.

The red box the proton signals coming from the amide backbone of the peptide. The purple box indicates the signals coming from the NH protons of Arg and Lys.

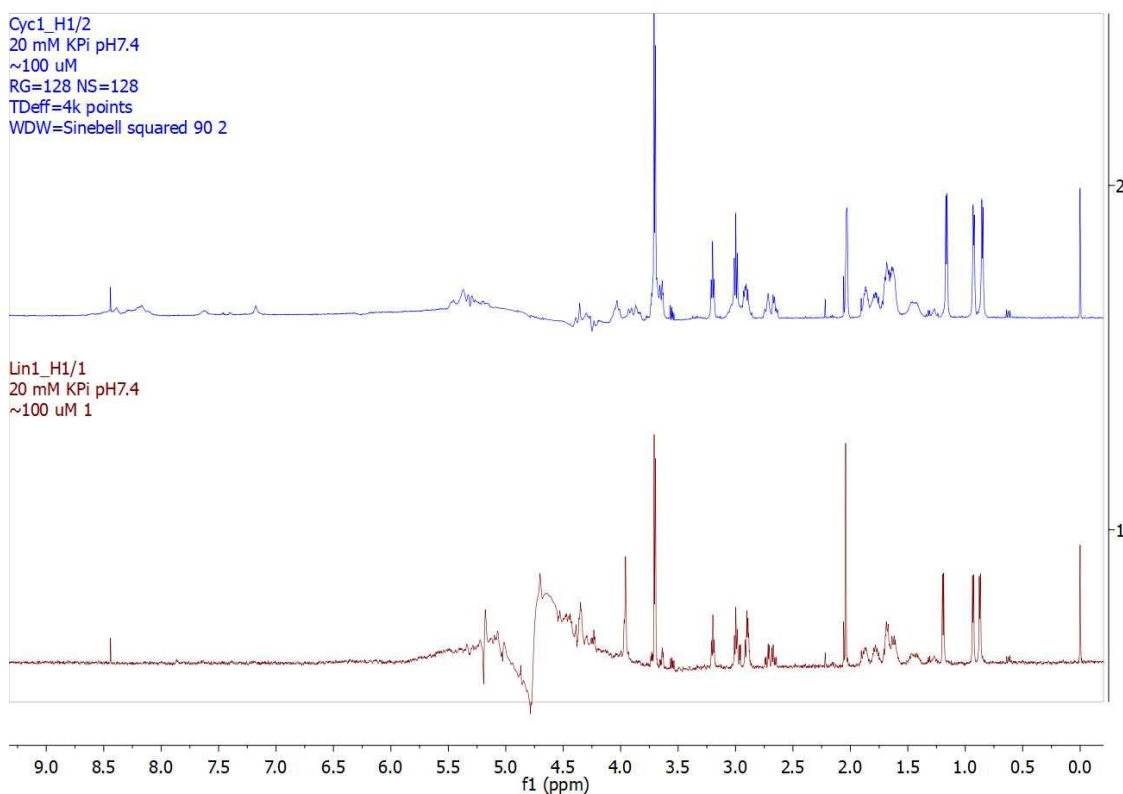


Figure 60  $^1\text{H}$ -NMR spectra of linear peptide 1 and loop 1 stacked over each other. Brown – the spectrum of the linear peptide 1. Blue – the spectrum of the loop 1.

The proton signals of the NH protons within amide bonds of loop 2 are more visible than in loop 1, however they are still relatively broad and clustered (**Figure 61**). This observation indicates, that loop 2 was more structured than loop 1. Moreover, the proton signals of the Asp NH<sub>2</sub> group and aromatic signals from Trp were observed in the spectrum. In the spectrum of loop 2 two populations of signals are present in the region of 3.2-3.4 ppm, which corresponds to the (<sup>β</sup>H) of Trp. This suggests that different conformers are present in the sample of this peptide.

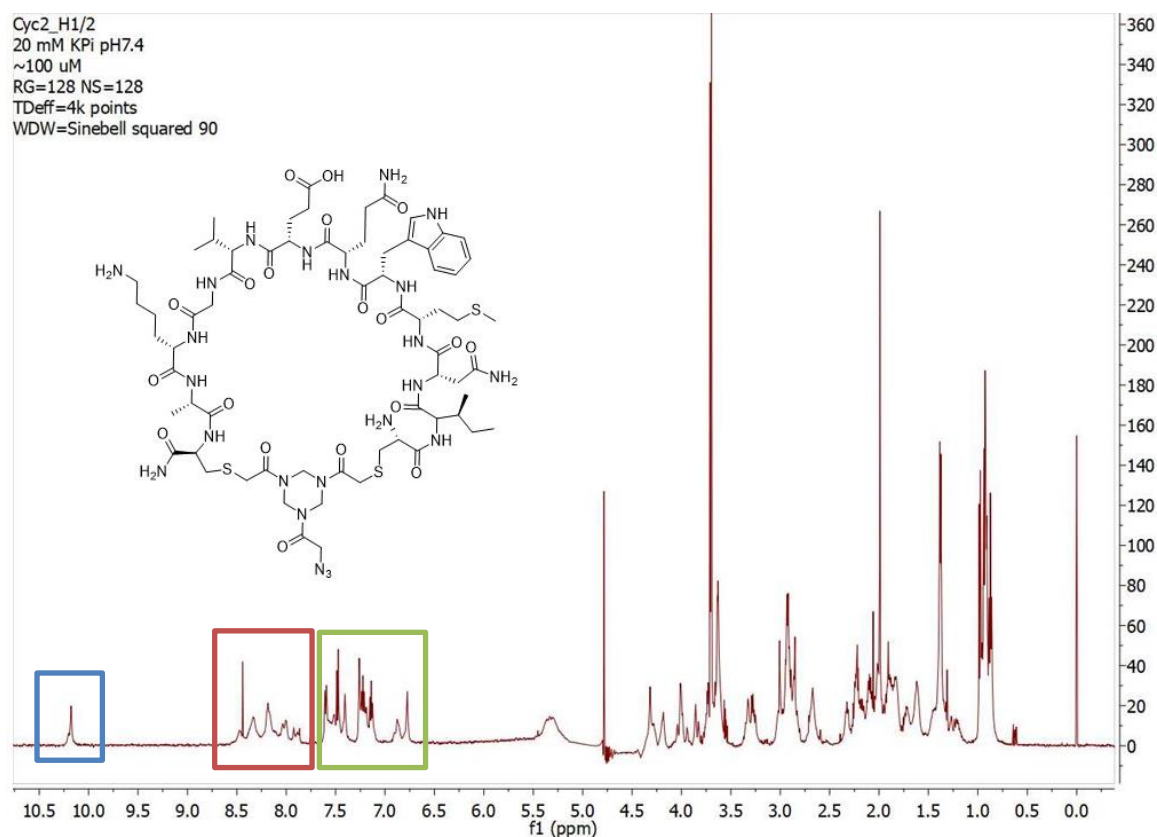


Figure 61 <sup>1</sup>H NMR spectra loop 2.

The red box indicates the proton signals coming from the amide backbone of the peptide. The green box indicates aromatic protons belonging to the indol ring of Trp and NH<sub>2</sub> protons of Asn. The blue box indicates the NH proton from the indol ring.

Once the two spectra were stacked over each other (**Figure 62**), the changes in the aliphatic region corresponding to the amino acid side chains become more visible. Since the hinge used for the cyclisation is in a chair-like conformation,<sup>[155]</sup> a conclusion can be drawn that the side chains of Asn, Gln, Trp and Lys might be interacting with each other causing the peptide to obtain a more constrained secondary structure. Linear peptide 2, due to precipitation, was at too low concentrations to be able to draw conclusions whether the amide backbone region was different from the loop 2.

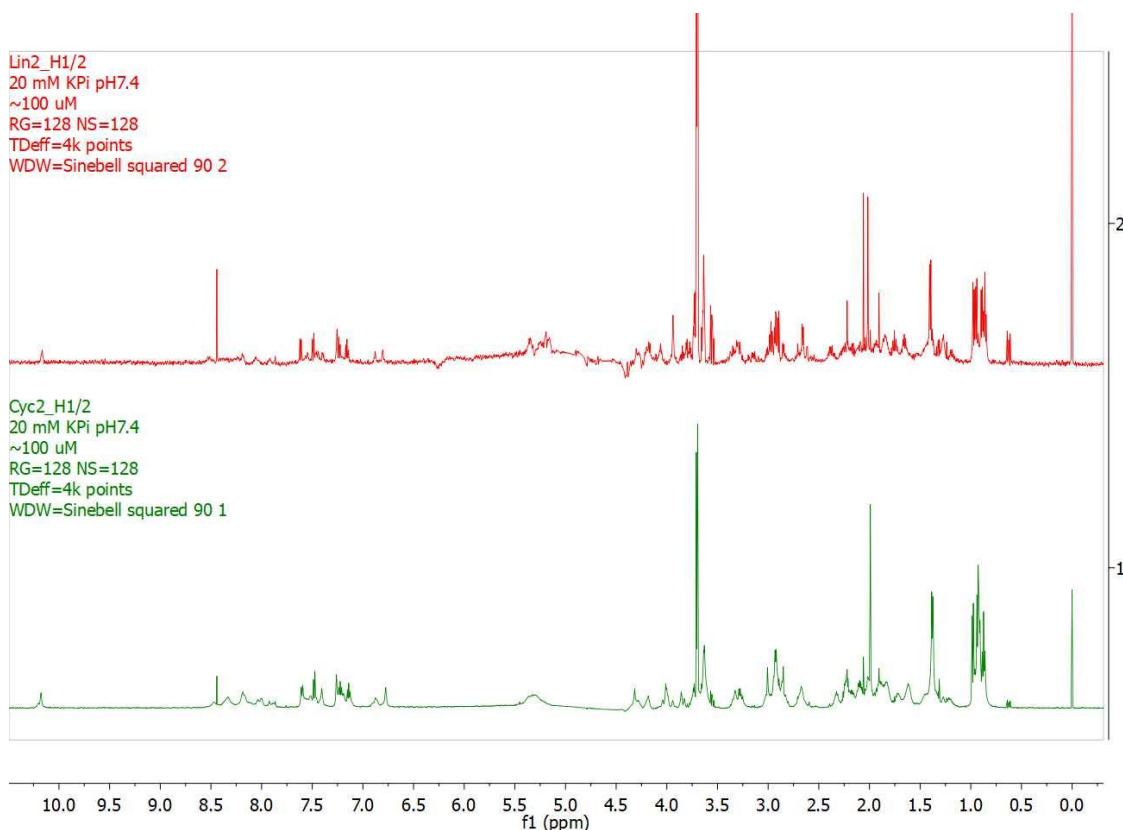


Figure 62  $^1\text{H}$  NMR spectra of linear peptide 2 and loop 2 stacked over each other. Red – the spectrum of the linear peptide 2. Green – the spectrum of the loop 2.

The proton signals present in the amide backbone region in the  $^1\text{H}$ -NMR spectrum of loop 3 are broader than those present in the spectra of loop 1 or 2 (**Figure 63**). Moreover, these signals are strong, which indicates that the amide backbone of the peptide is protected from the solvent because it is well constrained. As in the linear peptide, in the region 7.0-7.5 ppm signals from the  $\text{NH}_2$  protons of the C-terminus are visible.

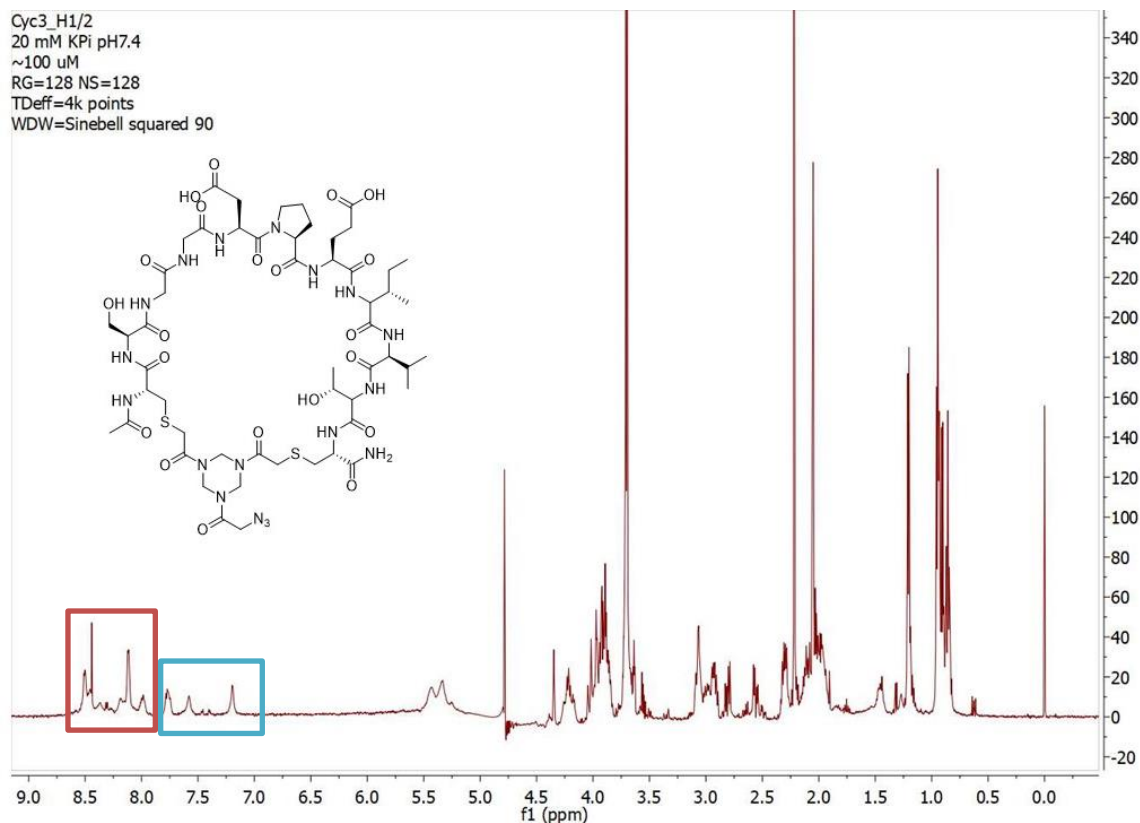


Figure 63  $^1\text{H}$  NMR spectra loop 3.

The red box indicates the proton signals coming from the amide backbone of the peptide. The cyan box indicates the protons signals coming from the C-terminal NHs.

After stacking of the  $^1\text{H}$ -NMR spectra of the linear peptide 3 and loop 3 over each other it became clearer, that better dispersed proton signals are present in the region of 7.0-7.5 ppm, corresponding to the  $\text{NH}_2$  protons of the C-terminus in the spectrum of loop 3 (**Figure 64**). Also changes in the chemical shifts of the protons in the amide backbone are present: in the spectrum of loop 3 these signals are broader and better resolved. These observations suggest, that the backbone of the loop 3 is constrained and ordered. More changes can be observed also in the region 0.8-1.0 ppm which corresponds to the proton signals of Val ( $\gamma\text{CH}_3$ ) and Ile ( $\delta\text{CH}_3$  and  $\gamma\text{CH}_3$ ). Also, very subtle changes can be observed at 1.15-1.25 ppm where  $\gamma\text{CH}_3$  protons of Thr are present. These changes might stem from the intramolecular hydrophobic interactions between the methyl groups of the aforementioned amino acids, which is another evidence towards the well-organised structure of the loop 3.

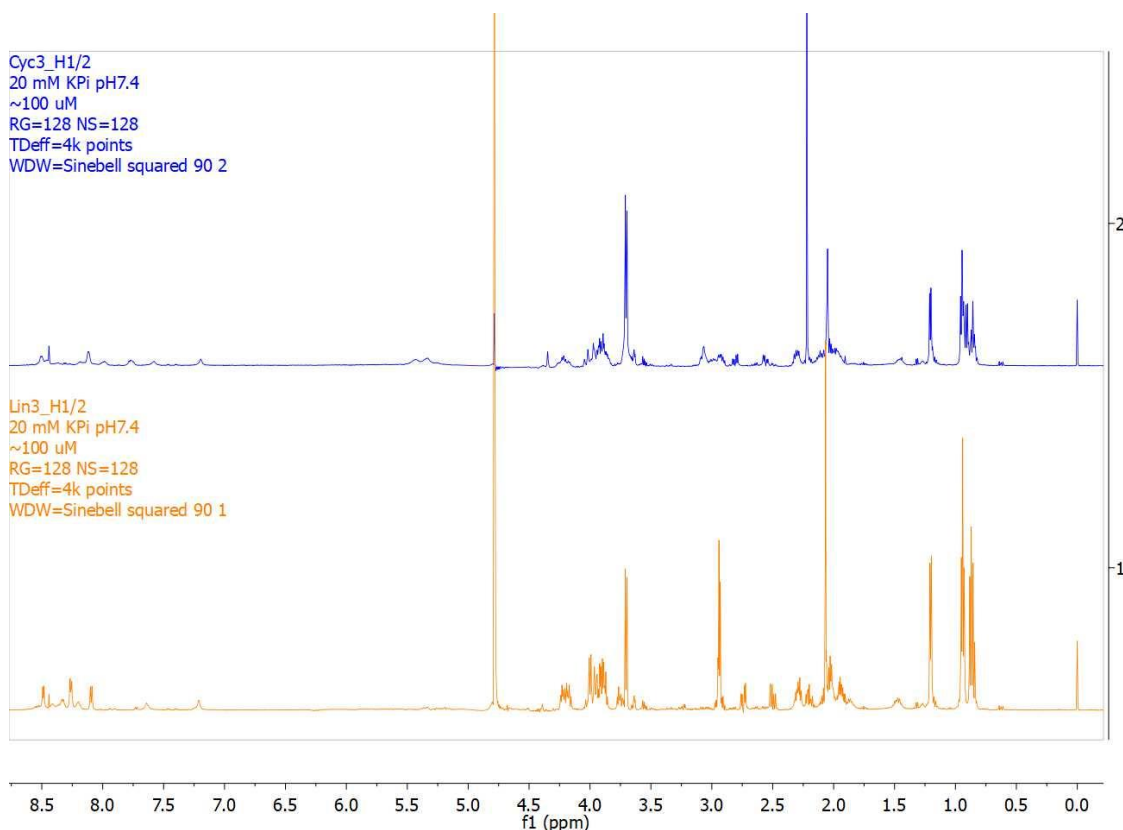


Figure 64  $^1\text{H}$  NMR spectra of linear peptide 3 and loop 3 stacked over each other. Orange – the spectrum of the linear peptide 3. Blue – the spectrum of the loop 3.

Overall, it can be concluded that out of three linear peptides, linear peptide 3 is the most constrained and linear peptide 1 is the most disordered. The same was observed among three loops.

### 3.3.1.3 gp120 mimics – scaffolded cyclic peptides

In order to find out whether the TAC scaffold induced further structural constraint and whether the peptide loops were interacting with each other once scaffolded,  $^1\text{H}$ -NMR spectra of the assembled peptides on the scaffold were obtained. If the loops were disordered and did not interact with each other on the scaffold, a spectrum which would be a sum of independent spectra would be expected. If the scaffold induced certain structural constraints, the chemical shift of the construct would differ from the signals of the independent loops. To monitor these changes, the spectra of constructs **22**, **26** and **30** were stacked over the spectra of loops 1-3. Moreover, the spectra of the three constructs were stacked over each other to compare whether there were any structural differences between them.

As an example, the spectrum of construct **22** stacked over loop 1, 2 and 3 is presented below in **Figure 65**. The different final concentrations of the samples did not allow to add up independent spectra. Therefore, it cannot be concluded, whether the spectrum of the construct **22** was or was not a simple sum of all the independent spectra. However, it can be observed that there are certain differences in the aliphatic region, especially at 0.7-1.1 ppm, which suggests that the loops might be interacting with each other once present on the scaffold. Similarly, the amide backbone and aromatic regions have a different overall shape.

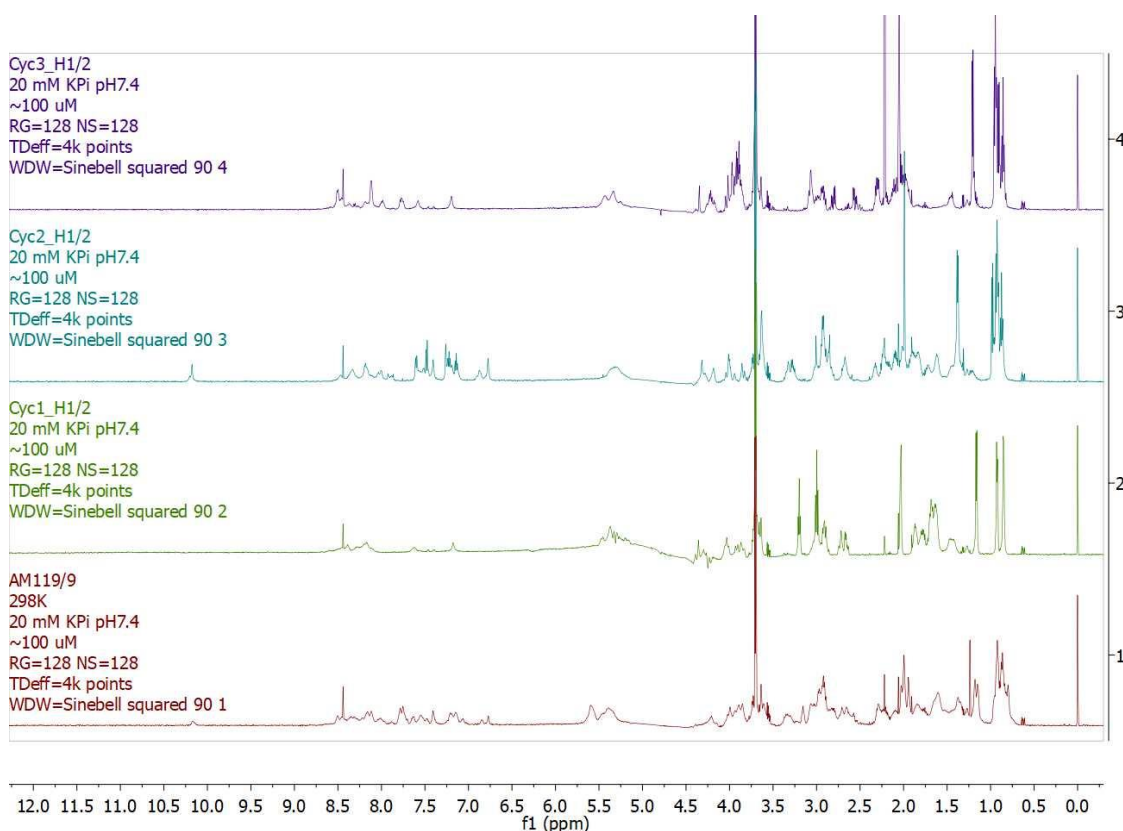


Figure 65 <sup>1</sup>H-NMR stacked spectra of construct **22**, loops 1, 2 and 3. Red: construct **22**, green: loop 1, cyan: loop 2, purple: loop 3.

The spectra of constructs **22** and **30** were stacked over each other, which revealed, that they were almost identical (**Figure 66**). However, very subtle changes in the region of 0.7-1.0 ppm corresponding to the methyl groups suggest that the amino acids containing these groups (Thr, Ala, Val, Ile, Leu, Met) might be interacting with each other differently once scaffolded.



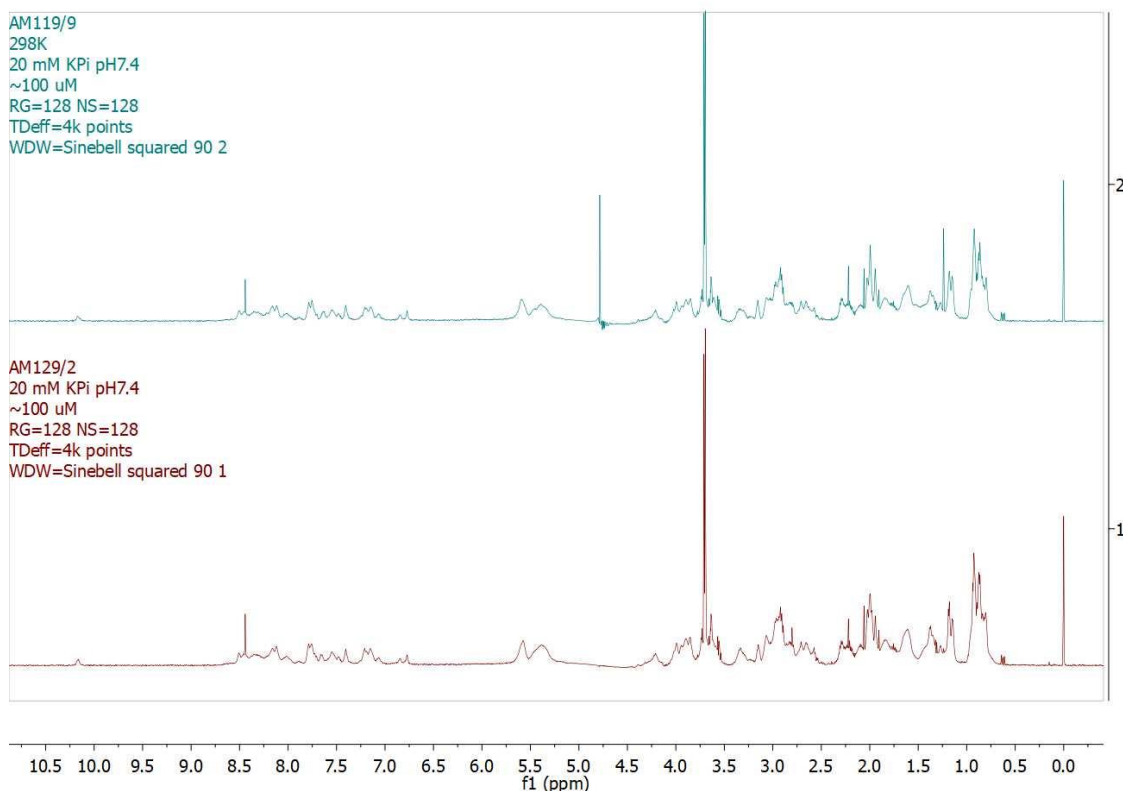


Figure 66  $^1\text{H}$ -NMR spectra of constructs **22** and **30** stacked over each other. Cyan: construct **22**, Red: construct **30**.

However, more subtle differences were observed once the  $^1\text{H}$ -NMR spectra of constructs **22** and **26** were stacked over each other (**Figure 67**). The peaks in the aromatic region of construct **26** are sharper than in the construct **22**. Also, the signal coming from the NH of the indol ring of Trp is sharper when compared to the corresponding signal in the construct **22**. There are two possible explanations for this observation. The first one assumes that the entire construct **26** obtains a very defined conformation and the peaks are sharp because there is no exchange with the surrounding solvent. The second hypothesis is that the construct **26** is less constrained than construct **22** (and construct **30**, since it is identical to construct **22**) and is present in different conformations. If the conformations of this construct are exchanging between each other quickly or the molecule in the sample is tumbling slowly, then the average signal becomes sharp. Due to the known cyclophane ring flipping<sup>[144]</sup> the second hypothesis is more probable.



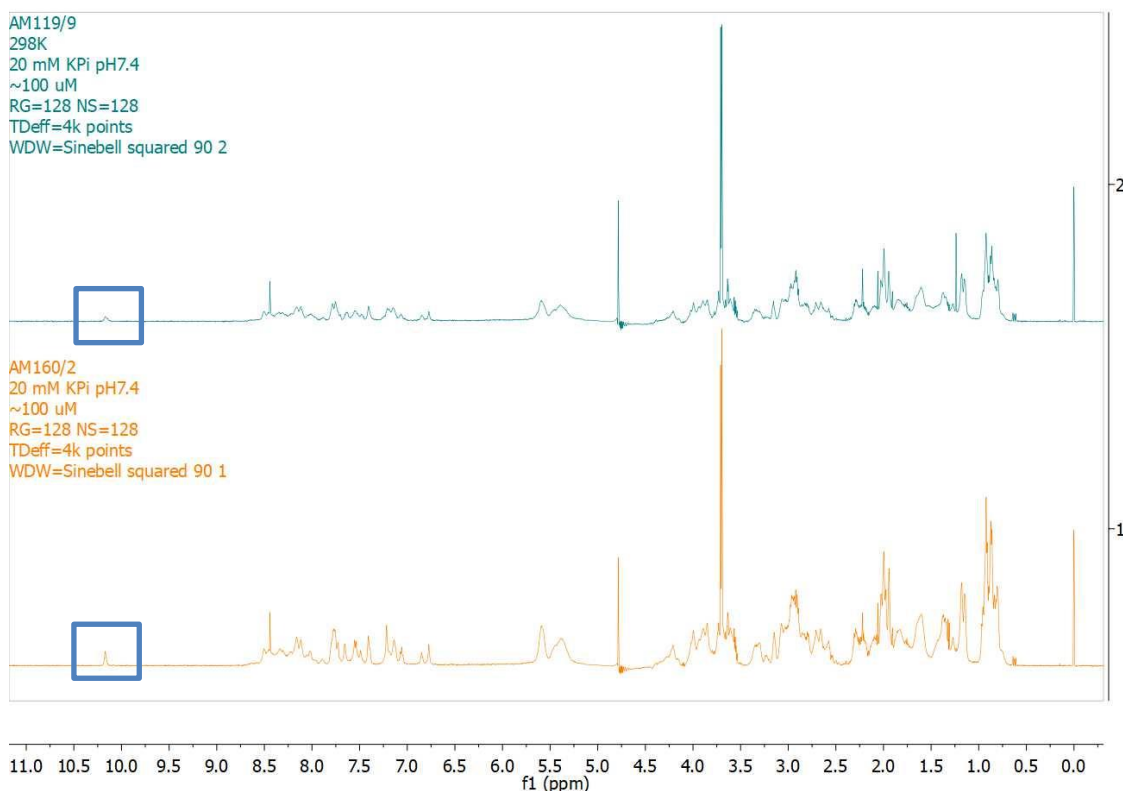


Figure 67  $^1\text{H}$  NMR spectra of constructs **22** and **26** stacked over each other. Cyan: construct **22**, orange: construct **26**. The blue box indicates the signal of the NH within the indol ring of Trp.

Moreover, construct **22** was subjected to  $^1\text{H}$ -NMR analysis at a range of temperatures to see whether the temperature change affects the obtained spectrum. The temperature was decreased by with steps of  $4^\circ$  from 298-278K (Figure 68). The obtained spectra were stacked over each other to investigate if any changes were apparent. The obtained spectra appear almost identical, with only very subtle peak sharpening in both aliphatic and amide-backbone regions.

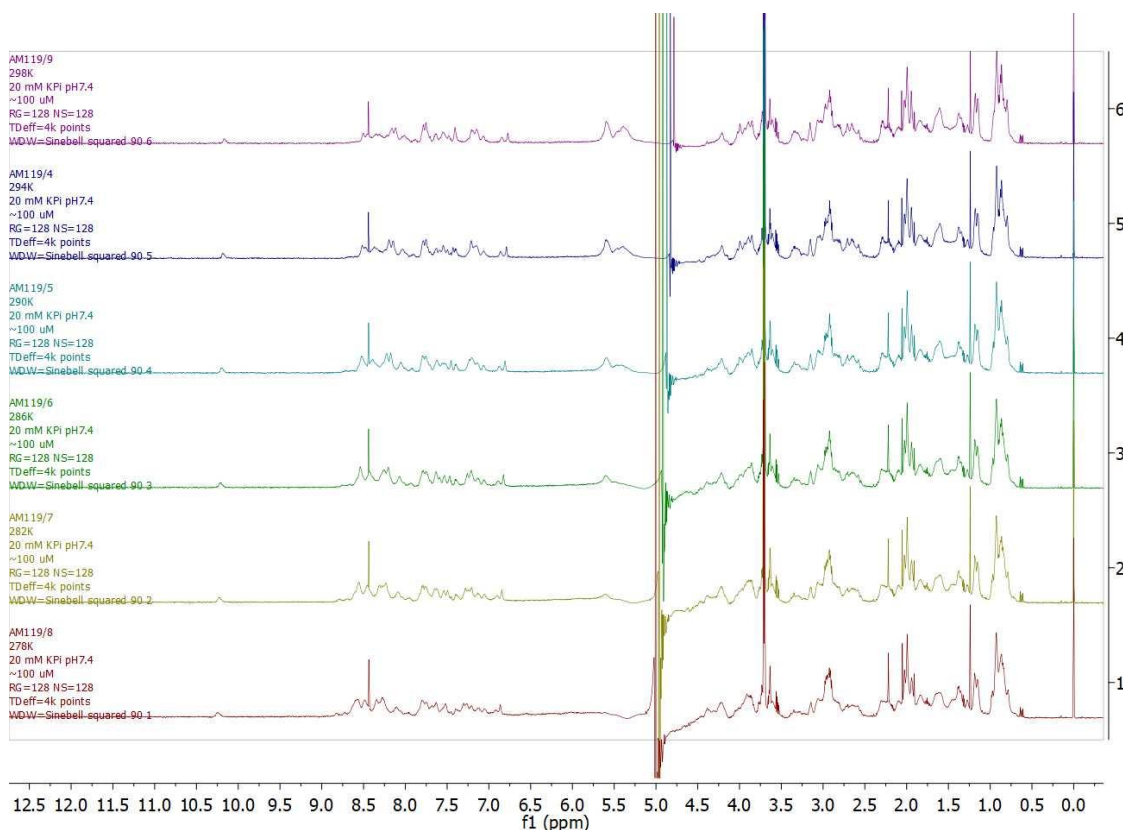


Figure 68 Variable temperature  $^1\text{H}$ -NMR of the construct **22**.  
The temperature was decreased by  $4^\circ$  in 5 steps from 298-278K.

### 3.3.2 CD - spectroscopy analysis of synthesized linear and cyclic peptides and assembled constructs

To complement the results obtained by  $^1\text{H}$ -NMR of the peptides and constructs, the same samples were used to obtain CD-spectroscopy spectra. The CD-spectroscopy experiments were performed and analysed with the assistance of Dr. Sharon Kelly (University of Glasgow, Institute of Molecular, Cell and Systems Biology). Linear peptide samples were analysed both with and without TCEP and minor differences were observed in the spectra.

The results of these experiments are summarised in **Table 9**.

Table 9 Summary of the secondary structure composition estimates – closest matching solution with all of the linear peptides, cyclic peptides and the constructs with assembled loops. The goodness of fits as judged by NRMSD. Values lower than 0.1 were favourable. Method used: Contin-LL (Provencher & Glockner method).<sup>[131]</sup>

Compound	Helix	Strand	Turns	Unordered	Total	NRMSD
Linear 1	0.096	0.297	0.250	0.367	1	0.090
Linear 1 +TCEP	0.094	0.327	0.244	0.335	1	0.136
Linear 2	0.060	0.404	0.121	0.416	1.001	0.249
Linear 2 + TCEP	0.061	0.396	0.123	0.421	1.001	0.190
Linear 3	0.117	0.282	0.248	0.353	1	0.030
Linear 3 + TCEP	0.093	0.304	0.242	0.360	0.999	0.039
Loop 1	0.086	0.334	0.236	0.344	1	0.143
Loop 2	0.070	0.367	0.126	0.437	1	0.261
Loop 3	0.084	0.330	0.233	0.353	1	0.080
Construct 22	0.067	0.372	0.126	0.434	0.999	0.136
Construct 26	0.073	0.359	0.127	0.440	0.999	0.131
Construct 30	0.071	0.373	0.124	0.432	1	0.138

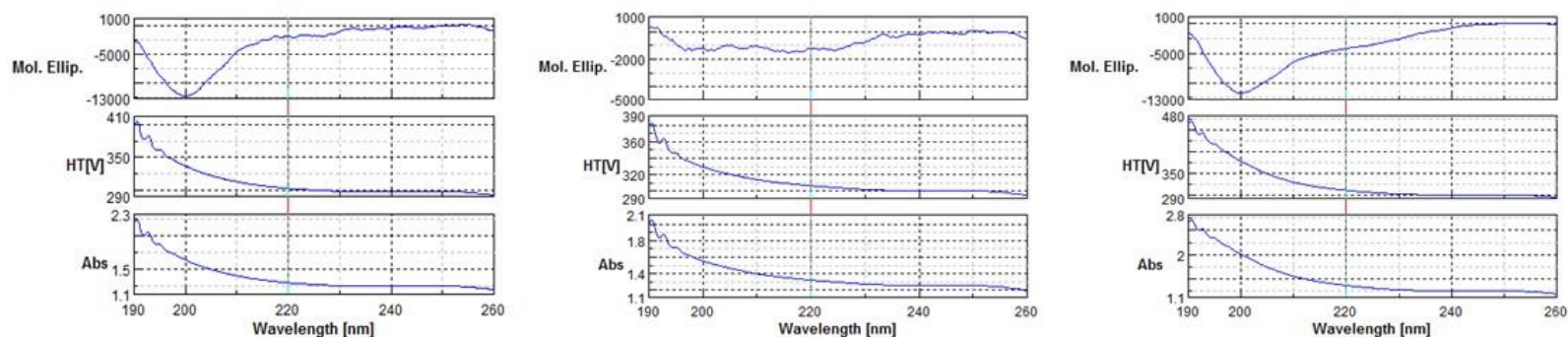
The samples in **Table 9** are described by the secondary structure composition and total content and the normalised root mean square deviation (NRMSD) value. NRMSD value represents a goodness-of-fit. A low NRMSD value ( $<0.1$ ) suggested that the analysis produced good results. Conversely, a high NRMSD value ( $>0.1$ ) suggested that the similarity between the calculated secondary structure and the actual structure (experimental) was unlikely to be correct.<sup>[157]</sup> As can be observed among the linear peptide samples, the addition of TCEP caused minor differences in the CD-spectra (the difference of maximum 3% between the values with and without TCEP). This suggested, that the linear peptides are present in the reduced form regardless of addition of the reducing agent. Comparison of the linear peptides with their cyclic counterparts allowed to observe that cyclisation of the peptides seemed to induce minor changes in the secondary structure composition.

Among the constructs with attached peptides loops, constructs **22** and **30** appear to be almost identical. Construct **26** appeared to be slightly more disordered. These observations were in agreement with conclusions drawn from  $^1\text{H}$  NMR experiments.

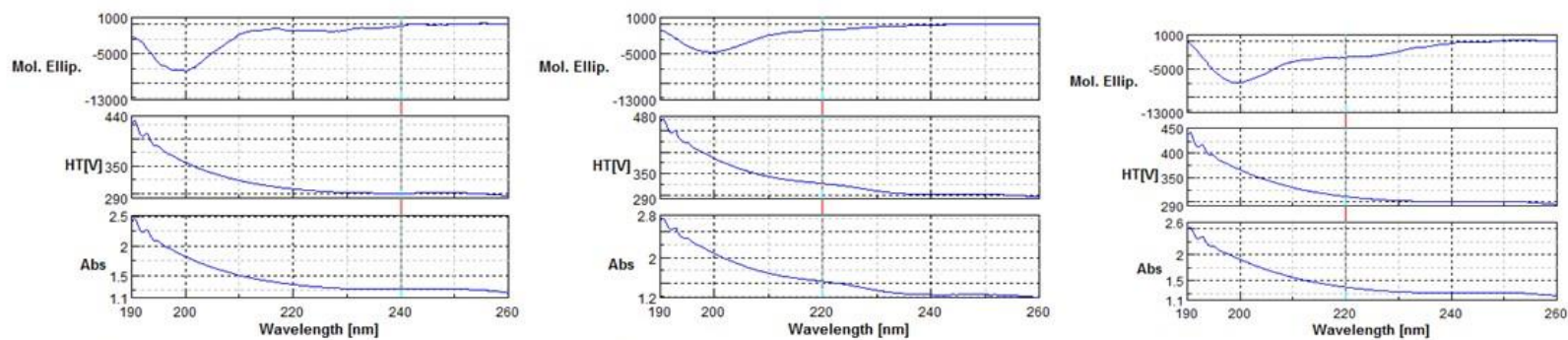
The obtained results suggest that all the obtained compounds are in 33-44% disordered. However, the NRMSD value for several compounds was unfavourable therefore conclusions drawn for these compounds were uncertain.

All synthesized compounds apart from linear peptide **2**, had the ellipticity minimum around 200 nm as shown in **Figure 69** (CD spectra for linear peptides are shown without TCEP. CD spectra of linear peptides supplemented with TCEP are in the Appendix, **Figure 164 - Figure 166**). Moreover, spectra of the peptides assembled on the TAC scaffold were superimposed on each other. As can be observed in **Figure 70**, compounds **22** and **30** are almost identical. This observation was in line with results obtained from  $^1\text{H}$  NMR experiments.

### Linear peptides



### Cyclic peptides



### Constructs

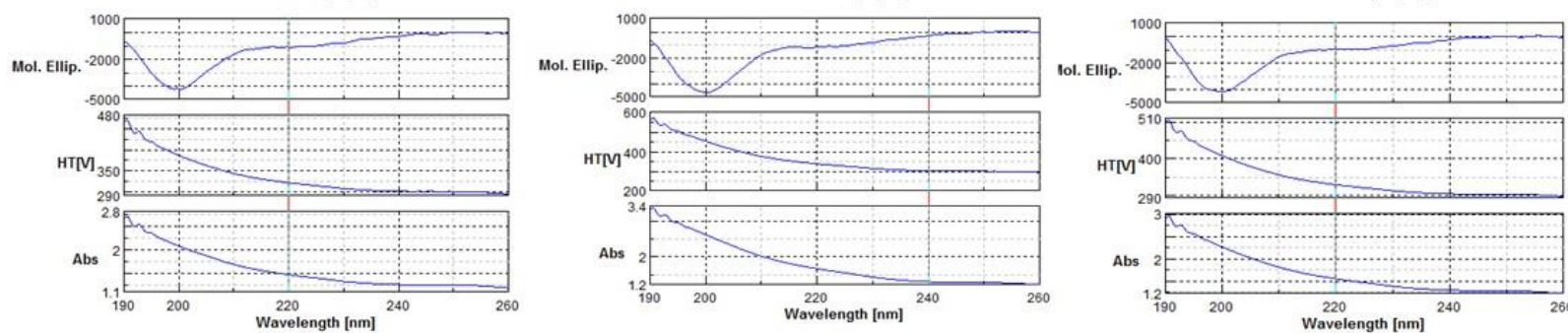


Figure 69 CD-spectra of the synthesized compounds.

In linear peptides (without TCEP) from left spectra of: linear peptide 1, 2 and 3. In cyclic peptides from left spectra of: loop 1, loop 2, and loop 3. In constructs from left spectra of: **22**, **26**, **30**.

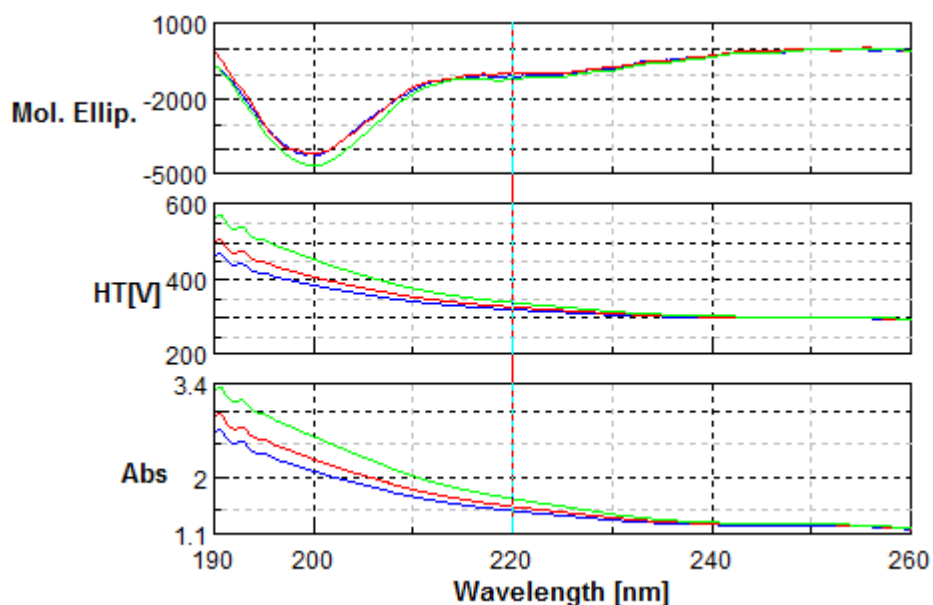


Figure 70 Superimposition of the CD-spectra of compounds **22** (blue), **26** (green) and **30** (red).

### 3.4 Conclusions

Three linear peptides corresponding to the discontinuous epitope sequences in gp120 were synthesized and cyclised with a novel cyclisation hinge to increase the solubility of the obtained cyclic peptides. These loops were then successfully sequentially attached to the TAC scaffold to obtain three different constructs (**22**, **26**, **30**) which would mimic the CD4-binding site of gp120. Moreover, the synthesis of the constructs was timewise shortened by attachment of two loops and TES-deprotection in one-pot reaction. An attempt to synthesize the full construct in one-pot was performed, however during the TIPS-deprotection the compound decomposed. All the obtained compounds were then analysed by  $^1\text{H}$ -NMR and CD-spectroscopy to gain further insight in to the secondary structure. From the obtained results it can be concluded that the linear peptide 3 is the most ordered, while the linear peptide 1 is the most disordered out of the three peptides. The same observation applies to the cyclic peptides, respectively. The analysis of the three constructs revealed, that two out of them are almost identical (**22** and **30**) and the third one (**26**) is possibly less constrained than the other two.

The biological evaluation of the obtained compounds by SPR will be described in the Chapter 4.

## 4 Evaluation of the binding of gp120 protein mimics to the CD4D12 by SPR

### 4.1 Introduction

The main objective of this project was to evaluate binding of the synthesized gp120 discontinuous epitope mimics to the CD4 receptor protein, which is the main port of entry of HIV-1. Previous attempts in the Liskamp group to evaluate binding of gp120 mimics to the CD4-receptor relied on an ELISA assay, which has proven to be unreliable since it often afforded variable results even with values below 0% or above 100% of control binding.<sup>[114]</sup> A robust, reproducible, and a reliable method to study binding properties of the synthesized gp120 mimics was needed. Initially it was decided, that binding between the mimics and the protein could be evaluated by ITC. However, ITC required large quantities of concentrated protein samples, therefore in order to meet these requirements, attempts towards CD4 and gp120 expression and purification were made. The obtained proteins were subjected to an ITC experiment; however, it was found that the concentrations of both proteins were too low to obtain any responses. The entire gp120 sample was consumed by this experiment and much more would be needed to obtain good quality results and to optimise the method. Moreover, large scale expression and purification of gp120 was cumbersome and not feasible, and in the end efforts in this direction were discontinued. Therefore, SPR was used since this technique requires smaller quantities and lower concentrations of the proteins. Moreover, in contrast to ELISA it should allow investigation of the kinetics and affinity of an interaction. In an SPR experiment, one of the molecules (ligand) is attached (covalently or non-covalently) to the sensor chip and a sample with the interacting partner (analyte) is passed over the surface (**Figure 71**).



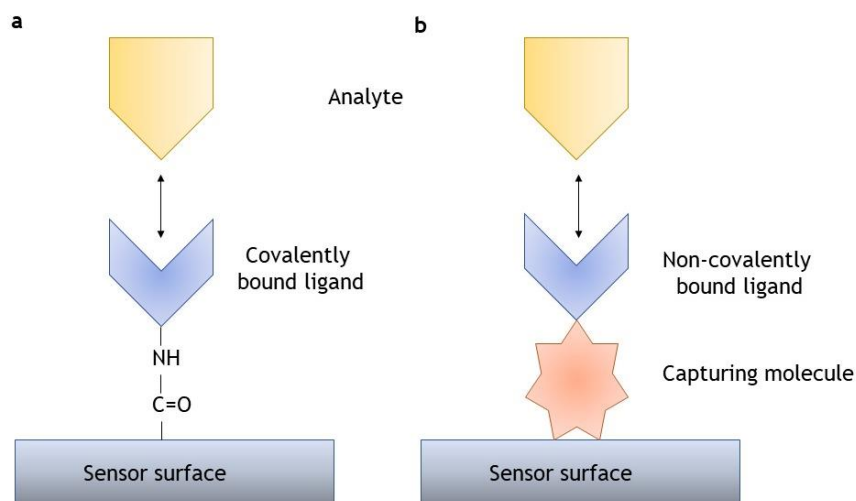


Figure 71 Illustration of the interacting partners in the SPR experiment.

a) A covalently immobilised ligand binds an analyte in the solution; b) A ligand is non-covalently bound to a capturing molecule (e.g. antibody, streptavidin) and interacts with an analyte in the solution.

Binding of the two molecules will generate a response, which is proportional to the bound mass on the surface and can be detected (**Figure 72**). In the literature, SPR has been successfully used for investigation of the binding between CD4 (both 4-domain and 2-domain) and gp120.<sup>[124][128][130][158][159]</sup> The principle of SPR is described in more detail in section 1.3.3.

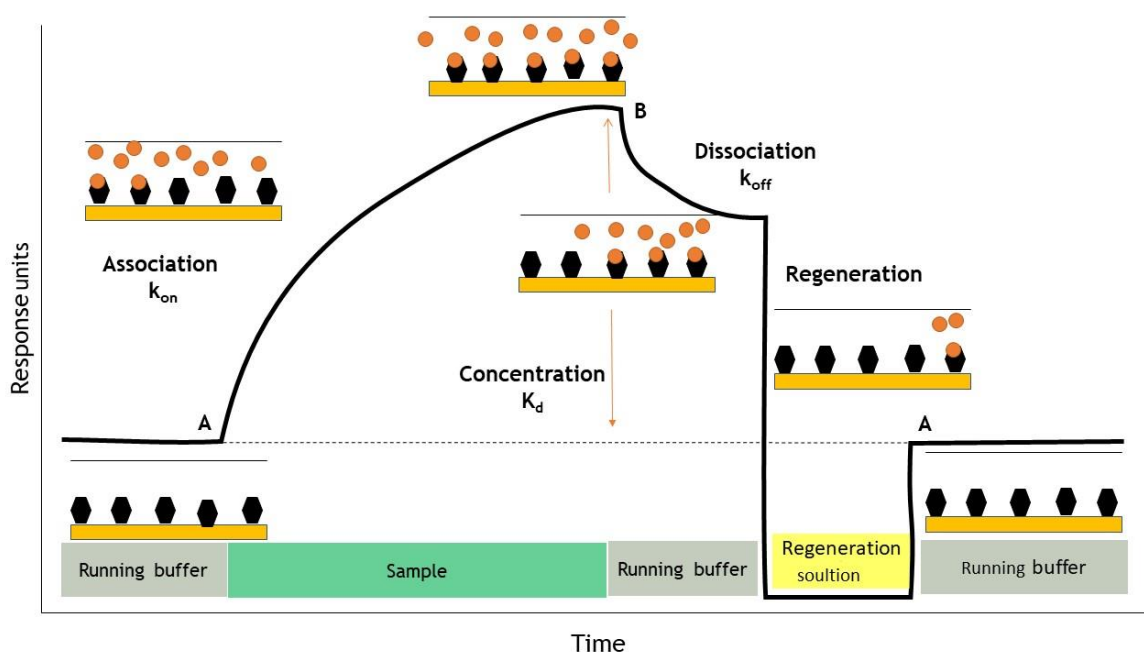


Figure 72 Response of the SPR experiment in the form of sensogram. Adapted from <sup>[160]</sup>.

CD4D12 and gp120 were obtained commercially. The two domain CD4D12 was obtained to set up an experiment and compare the activity and binding profile of



the expressed CD4D12 with the commercial product. Binding of gp120 to CD4D12 served as a positive control that could be compared with the literature data.

## 4.2 Results and discussion

An SPR experiment could be set up in two ways: as a direct binding assay, where binding of gp120 mimics to immobilised CD4D12 is measured or as a competition assay, where inhibition of the binding of CD4D12 to immobilised gp120 in presence of gp120 mimics is measured. Both approaches were investigated in order to find optimal conditions for evaluation of binding properties of synthesized gp120 mimics, as described in sections 4.2.1 and 4.2.2. However, the first objective was to optimise experimental conditions of gp120-CD4D12 binding as a positive control, as the obtained results could be compared with the literature data. Only then, binding of gp120 mimics to CD4D12 could be measured.

The process of setting up an SPR experiment can be divided into four phases: 1) determination of the optimal pH for immobilisation (pH scouting) and pre-concentration, 2) immobilisation of the ligand, 3) determination of regeneration conditions (regeneration scouting), and 4) the binding experiment.

The CM5 sensor chip (Biacore) was chosen for the SPR experiments as it was described in the literature.<sup>[124][128][130][158][159]</sup>

### 4.2.1 Towards a competition assay of the binding of CD4D12 to the immobilised gp120

In an SPR experiment, the sample that is the analyte and passed over the sensor surface is needed in larger quantities than the sample of an immobilised ligand. Gp120 had to be commercially obtained, while CD4D12 could be expressed and purified in larger quantities, therefore once the experimental conditions were determined using commercial proteins, in house expressed CD4D12 could be easily used. To minimise utilisation of gp120 protein we decided to perform a competition assay. To achieve this, gp120 had to be immobilised on the sensor surface while CD4D12 was passed over it and the binding was measured to obtain a positive control. In the competition experiment, gp120 mimics would be pre-incubated with the CD4D12 protein and passed over immobilised gp120. The decrease in binding of CD4D12 to gp120 in relation to gp120-CD4D12 binding in the

absence of gp120 mimics would allow an estimation of the binding capacity of the mimics.

#### 4.2.1.1 pH scouting and pre-concentration

In order to achieve the desired immobilisation level of the highly diluted (typically  $10\text{--}50\ \mu\text{g}\cdot\text{mL}^{-1}$ ) protein on the sensor surface, it had to be electrostatically pre-concentrated in the negatively charged dextran matrix of the sensor. The carboxymethylated dextran on the sensor surface is negatively charged at a pH above 3.5, therefore to efficiently concentrate ligand on the surface it has to obtain a positive charge. To optimally pre-concentrate the ligand on the sensor surface, the ligand should be present in a solution at a pH between 3.5 and its isoelectric point (pI). Thus, the ligand and the sensor surface carry opposite net charges.

To find optimal coupling conditions for gp120 immobilisation, gp120 under different buffer conditions was passed over an unactivated sensor chip (pH scouting) and the extent of its preconcentration was observed as an increase in response.

Gp120 was prepared in 10 mM acetate buffers at different pH values: 5.5, 5.0 and 4.5 and exposed to the unactivated surface of a CM5 chip. After each protein injection, 1 M ethanolamine solution at pH 8.5 was injected to remove all electrostatically bound ligand. As is shown in **Figure 73** pre-concentration levels of gp120 on the chip surface vary at different pH values. At pH 5.5 the lowest pre-concentration level was obtained (6127.6 response units, (RU)). At pH 5.0 the pre-concentration level was the highest (7215.6 RU). At pH 4.5 the pre-concentration level was also reasonably high (6913.6 RU), however it could be observed that the bound protein dissociated to a lesser extent after injection. Therefore, lower pH values were not tested, as they could lead to protein aggregation or denaturation. The optimal pH for immobilisation was determined to be 5.0.

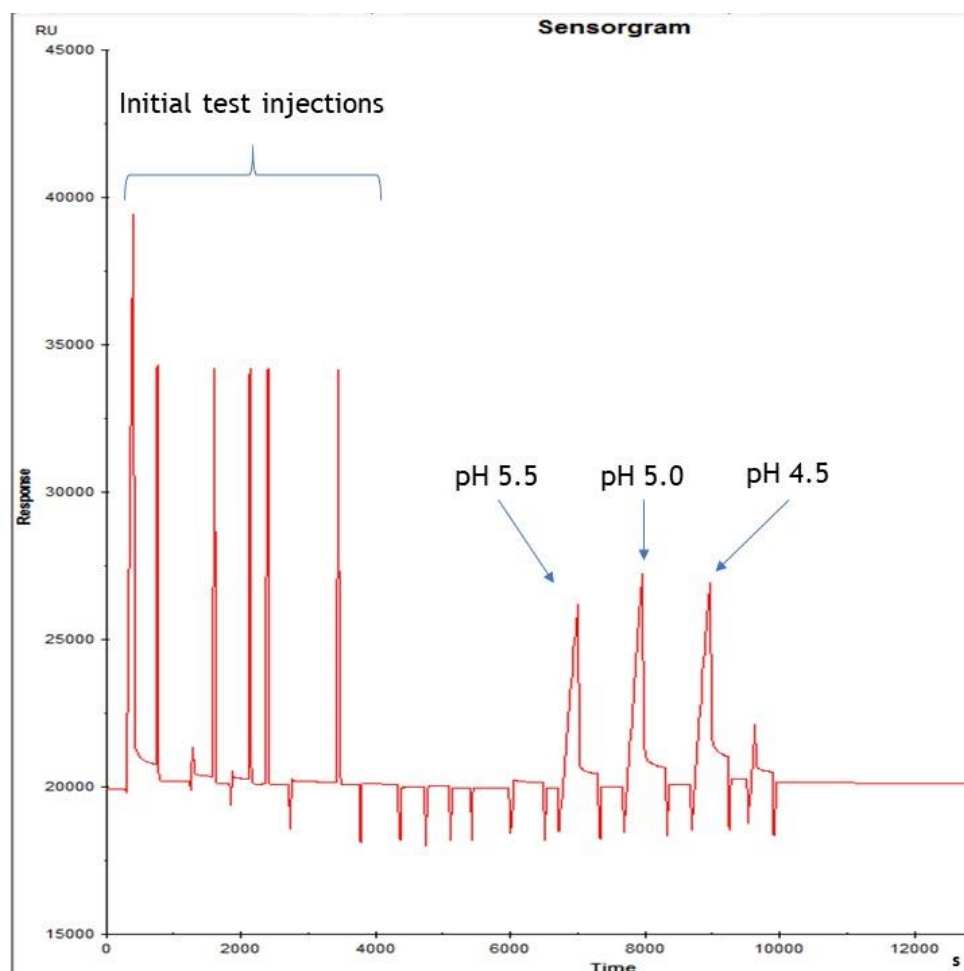


Figure 73 pH scouting and pre-concentration studies of gp120. gp120 ( $10 \mu\text{g}\cdot\text{mL}^{-1}$ ) in 10 mM acetate buffers at pHs: 5.5, 5.0, 4.5 injections over unactivated CM5 chip over 300 s.

#### 4.2.1.2 Immobilisation of gp120

The approach used for gp120 immobilisation was the covalent attachment of the protein to the surface of the sensor chip.

To immobilise gp120 on CM5 chip, firstly, the sensor surface in the flow cell 2 was activated by preparation of the hydroxy succinic ester derivatives of the carboxylic acid moieties on the dextrane layer by injection of 0.4 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC)/0.1 M N-hydroxysuccinimide (NHS) solution. Activation was followed by injection of a gp120 solution. In the first immobilisation test, gp120 was present at a concentration of  $10 \mu\text{g}\cdot\text{mL}^{-1}$  and was injected over 105 s. The last step was removal of the non-covalently bound ligand and deactivation of the chip surface by conversion of the remaining active ester groups to amides by injection of 1 M ethanolamine pH 8.5. The amount of immobilised gp120 was 1900 RU (**Figure 74**).

Later on, gp120 was used at  $20\ \mu\text{g}\cdot\text{mL}^{-1}$  was injected in short pulses until the desired amount of the protein on the surface of the chip was achieved (3600 RU or 1000 RU).

The test chip with 1900 RU of immobilised gp120 was used to find optimal regeneration conditions. The chip with 3600 RU of immobilised gp120 was used to investigate binding of CD4D12 to gp120.

Flow cell 1 served as a reference surface, therefore it was activated and deactivated similarly to the active surface in flow cell 2, albeit not exposed to gp120.

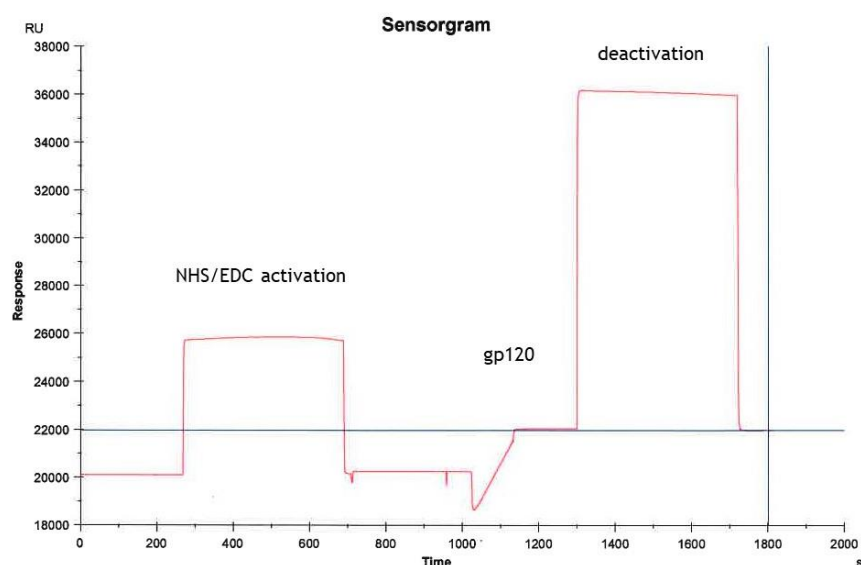


Figure 74 The process of gp120 covalent immobilisation on the surface of CM5 chip.

#### 4.2.1.3 Regeneration scouting

Regeneration of the chip consisted of removal of the bound analyte to prepare the chip surface for the next analysis cycle, ideally without destroying the activity of the immobilised protein. Efficient regeneration allowed the reuse of the chip in the next assays, therefore finding the optimal conditions was of the utmost importance. Determination of the regeneration conditions was achieved by evaluation of the different regeneration conditions by repeated injections of the analyte at the highest concentration planned for experiments and subsequent injection of a tested regeneration buffer. After each regeneration cycle response levels were examined to find the optimal regeneration procedure.

Since several SPR experimental procedures were described for the gp120-CD4 interaction, the procedure described by Cerutti et al.<sup>[130]</sup> was chosen where 10 mM glycine at pH 2.5 was used as a starting point in determining the regeneration procedure.

Three regeneration buffers were tested, these contained: 10 mM glycine at different pH values: 2.5, 2.0 and 1.5 (**Figure 75**). The regeneration was started by testing the mildest regeneration conditions (pH 2.5). At pH 2.5 an increase in baseline was observed, which suggested that the analyte was not removed completely and accumulated on the sensor surface. At pH 2.0 the baseline was still higher than the sample response, while at pH 1.5 both baseline and sample response started decreasing, which suggested deterioration of the analyte binding capacity. Therefore, intermediate conditions of 10 mM glycine at pH 1.75 were chosen as the regeneration buffer conditions.

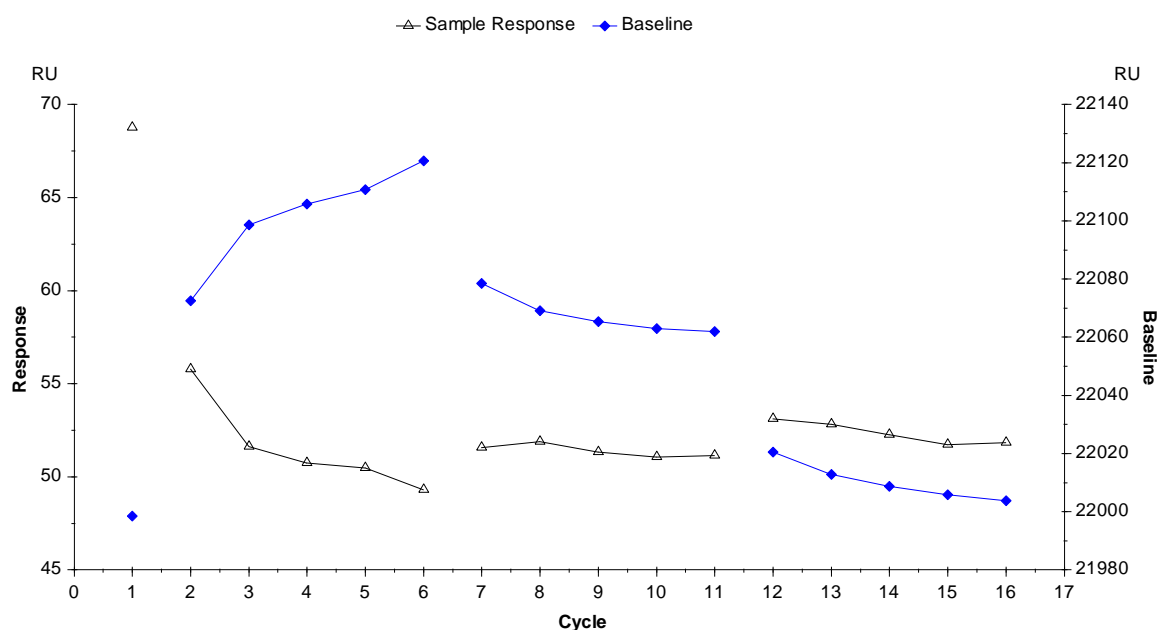


Figure 75 Regeneration scouting.

The ligand CD4D12 was immobilised and gp120 was injected over it at the concentration of 100 nM. The conditions tested from the left: 10 mM glycine pH 2.5, 10 mM glycine pH 2.0, 10 mM glycine pH 1.5.

#### 4.2.1.4 Binding of CD4D12 to immobilised gp120

Serial dilutions (0-250 nM) of CD4D12 were prepared in 1×PBS-P buffer and injected over 3600 RU immobilised gp120 on the CM5 chip. The association time was 120 s, followed by 600 s of dissociation and finally removal by injection of 10 mM glycine at pH 1.75 for 60 s. Each concentration of CD4D12 was injected

once, starting from the lowest to the highest concentration. A 15.6 nM CD4D12 solution was injected for a second time after all the samples were tested, to monitor if the response is at the same level after the entire assay. This experiment was repeated twice. Binding of the CD4D12 to gp120 was fitted a 1:1 Langmuir binding model describing the interaction of the two molecules in a 1:1 complex.

It was observed however, that the binding model did not fit the curve corresponding to the injection of the highest concentration (250 nM) of CD4D12. Therefore, this concentration was excluded from the kinetic analysis. The sensograms corresponding to the 3 replicates of the experiment are shown in **Figure 76**, **Figure 77** and **Figure 78**.

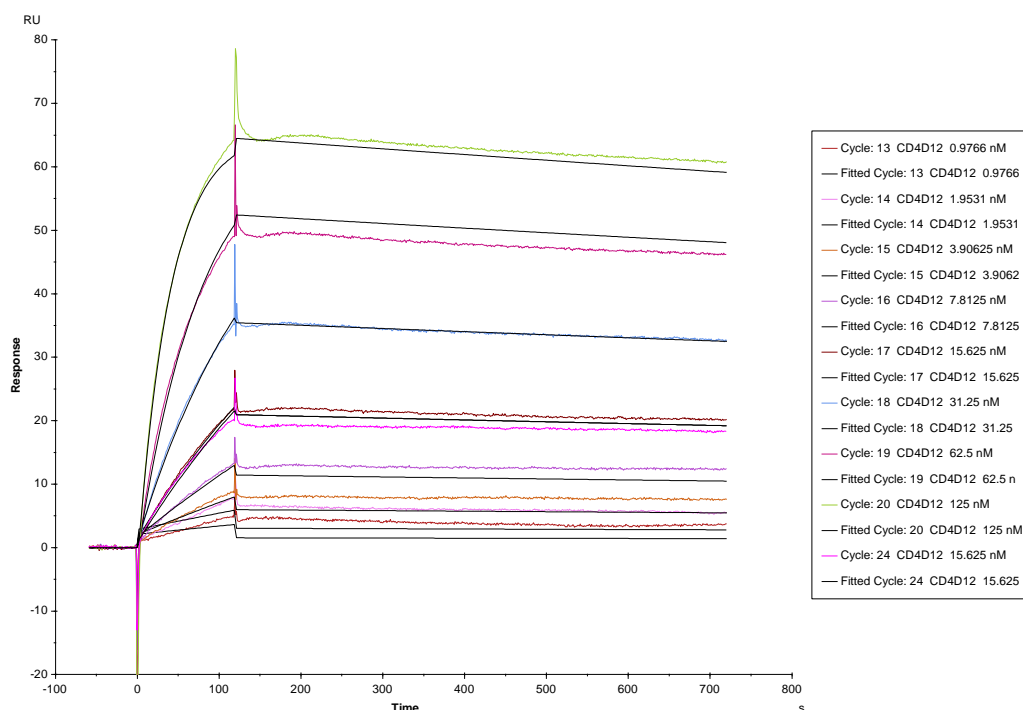


Figure 76 Replicate 1 of the interaction of analyte CD4D12 with immobilised ligand gp120 at 25°C. The coloured lines depict the double-referenced sensograms obtained from injections of 0, 125, 62.5, 31.3, 15.6, 7.8, 3.9, 1.9, 0.9 nM analyte CD4D12 across 3600 RU immobilised ligand gp120. Injection of 15.6 nM CD4D12 was repeated at the end of the experiment. The ligand surface was regenerated with a 60 s injection of 10 mM glycine pH 1.75 between each binding cycle. The black lines depict the global fit of the data to a 1:1 Langmuir interaction model, yielding  $k_a = 1.97 \cdot 10^5 \text{ M}^{-1} \text{ s}^{-1}$ ,  $k_d = 1.45 \cdot 10^{-4} \text{ s}^{-1}$ ,  $K_d = 7.36 \cdot 10^{-10} \text{ M}$ .

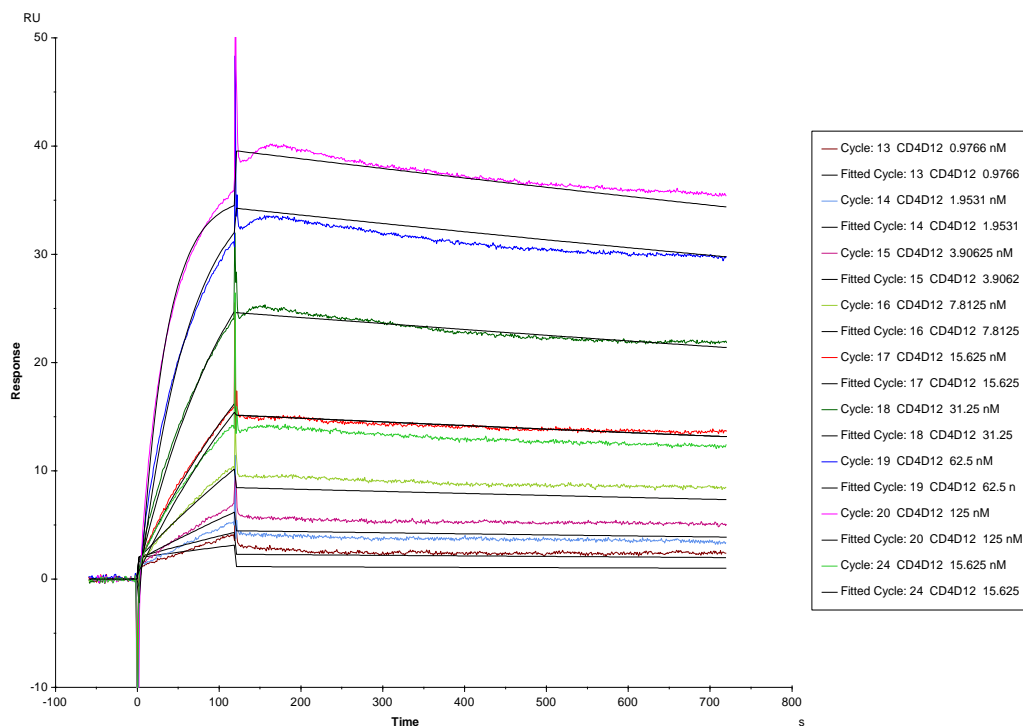


Figure 77 Replicate 2 of the interaction of analyte CD4D12 with immobilised ligand gp120 at 25°C. The coloured lines depict the double-referenced sensograms obtained from injections of 0, 125, 62.5, 31.3, 15.6, 7.8, 3.9, 1.9, 0.9 nM analyte CD4D12 across 3600 RU immobilised ligand gp120. Injection of 15.6 nM CD4D12 was repeated at the end of the experiment. The ligand surface was regenerated with a 60 s injection of 10 mM glycine pH 1.75 between each binding cycle. The black lines depict the global fit of the data to a 1:1 Langmuir interaction model, yielding  $k_a = 2.52 \cdot 10^5 \text{ M}^{-1} \text{ s}^{-1}$ ,  $k_d = 2.34 \cdot 10^{-4} \text{ s}^{-1}$ ,  $K_d = 9.30 \cdot 10^{-10} \text{ M}$ .

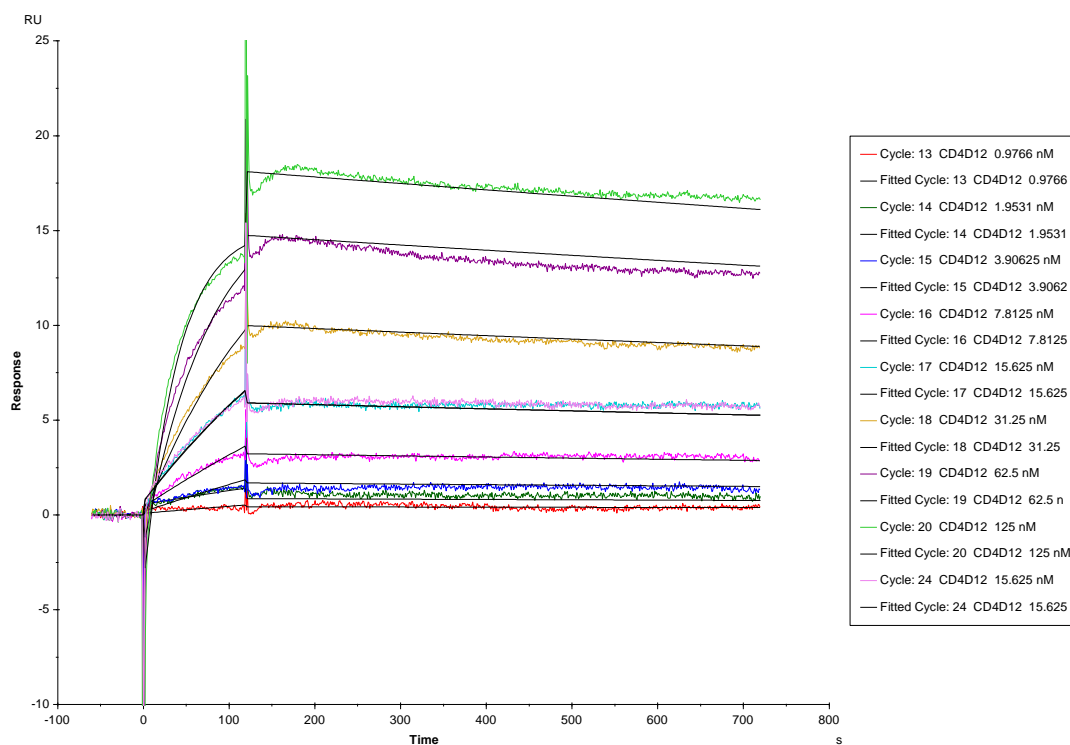


Figure 78 Replicate 3 of the interaction of analyte CD4D12 with immobilised ligand gp120 at 25°C. The coloured lines depict the double-referenced sensograms obtained from injections of 0, 125, 62.5, 31.3, 15.6, 7.8, 3.9, 1.9, 0.9 nM analyte CD4D12 across 3600 RU immobilised ligand gp120. Injection of 15.6 nM CD4D12 was repeated at the end of the experiment. The ligand surface was regenerated with a 60 s injection of 10 mM glycine pH 1.75 between each binding cycle. The black lines depict the global fit of the data to a 1:1 Langmuir interaction model, yielding  $k_a = 2.52 \cdot 10^5 \text{ M}^{-1} \text{ s}^{-1}$ ,  $k_d = 2.34 \cdot 10^{-4} \text{ s}^{-1}$ ,  $K_d = 9.30 \cdot 10^{-10} \text{ M}$ .

The obtained kinetic data are summarised in **Table 10**.

Table 10 Kinetic data obtained from the three experiments of binding CD4D12 to immobilised gp120.  $k_a$  – association rate,  $k_d$  – dissociation rate,  $K_d$  – equilibrium dissociation constant,  $R_{max}$  – analyte binding capacity.

Replicate	$k_a$ ( $M^{-1} s^{-1}$ )	$k_d$ ( $s^{-1}$ )	$K_d$ (M)	$R_{max}$ (RU)
1	$1.97 \cdot 10^5$	$1.45 \cdot 10^{-4}$	$7.36 \cdot 10^{-10}$	68.39
2	$2.52 \cdot 10^5$	$2.34 \cdot 10^{-4}$	$9.27 \cdot 10^{-10}$	40.78
3	$1.99 \cdot 10^5$	$1.95 \cdot 10^{-4}$	$9.81 \cdot 10^{-10}$	19.2
Mean	$2.16 \cdot 10^5$	$1.91 \cdot 10^{-4}$	$8.81 \cdot 10^{-10}$	-
Standard deviation	15%	23%	15%	-

After analysis of the obtained kinetic data (**Table 10**), it was observed that the activity of gp120 immobilised was decreasing, which was obvious from the decreasing  $R_{max}$  value. As can be observed in **Figure 76**, the fitted model overlayed with the obtained data set, however in the next two experiments (**Figure 77** and **Figure 78**) the models did not fit the highest concentrations any longer. The selected obtained responses from the three experiments at different concentrations were compared (**Table 11**). One cycle can be defined as an injection of the analyte at a certain concentration, followed by regeneration of the chip surface. As can be observed, within replicate 1 the binding level response between cycles 17 and 24 decreased by 1.2 RU, in replicate 2 by 1.4 RU and in replicate 3 by 0.1 RU. Between the cycle 17 of replicate 1 and 2 there are 24 cycles and the binding level between these decreased by 6.3 RU. From these observations it was calculated, that the response decreased by 0.013-0.016 RU per cycle. In the third replicate the responses were 4× lower when compared to the first replicate. This decrease in binding level was caused by degradation of gp120 by regeneration and over time. The decrease in gp120 activity was connected to the decrease in the concentration of active gp120, therefore decrease in the  $K_d$  value was observed (**Table 10**).



Table 11 Comparison of the obtained binding level responses (RU) at different concentrations of CD4D12 in the consequent experiments.

Cycle	Concentration (nM)	Binding level (RU)		
		Replicate 1	Replicate 2	Replicate 3
17	15.6	20.1	13.8	5.2
19	62.5	46.8	29.1	10.8
20	125	62.2	33.9	12.7
24	15.6 (second injection)	18.9	12.4	5.1

The previously published kinetic constants for interaction between CD4D12-gp120 were as follows:  $K_d = 6\text{-}52\text{ nM}$ ,  $k_a = 5.3\cdot 10^3\text{-}4.3\cdot 10^4\text{ M}^{-1}\cdot\text{s}^{-1}$ , and  $k_d = 2.6\cdot 10^{-4}\cdot\text{s}^{-1}$ .<sup>[128][124][130]</sup> Here a  $K_d$  of  $0.7\text{ nM}$ ,  $k_a = 2.28\cdot 10^5\text{ M}^{-1}\cdot\text{s}^{-1}$  and  $k_d = 1.68\cdot 10^{-4}\cdot\text{s}^{-1}$  (mean values) were obtained.

The equilibrium dissociation constant  $K_d$  is described by the equation:  $K_d = \frac{[A][B]}{[AB]}$ , where  $[A]$  and  $[B]$  are concentrations of interacting partners and  $[AB]$  is the concentration of the AB complex. The lower the  $K_d$ , the higher the stability of the formed complex. However, if the concentration of one of the interacting partners decreases, the  $K_d$  also decreases. Therefore, the obtained  $K_d$  of  $0.7\text{ nM}$  might be caused by the decreased concentration of the immobilised active gp120.

Once the positive control was obtained by testing both commercial proteins, the experiment with the expressed and purified CD4D12 was performed. As it was shown in the experiments with commercial CD4D12, at the concentration of  $250\text{ nM}$  the models did not fit the data set, it was decided to decrease the concentration of tested CD4D12 down to  $100\text{ nM}$ .

Serial dilutions of the expressed CD4D12<sub>ex</sub> were prepared ( $0\text{-}100\text{ nM}$ ) in  $1\times\text{PBS-P}$  buffer and injected over  $1000\text{ RU}$  of immobilised gp120 on the CM5 chip. The association time was  $120\text{ s}$ , followed by  $600\text{ s}$  of dissociation and finally removal

by injection of 10 mM glycine at pH 1.75 for 60 s. Each concentration of CD4D12 was performed in duplicate, starting injections from the lowest to the highest. The experiment was carried out twice. Binding of the CD4D12 to gp120 was fitted a 1:1 Langmuir binding model describing the interaction of two molecules in a 1:1 complex.

The sensogram of one of the two experiments is shown on **Figure 79** (for the second replicate see Appendix **Figure 167**).

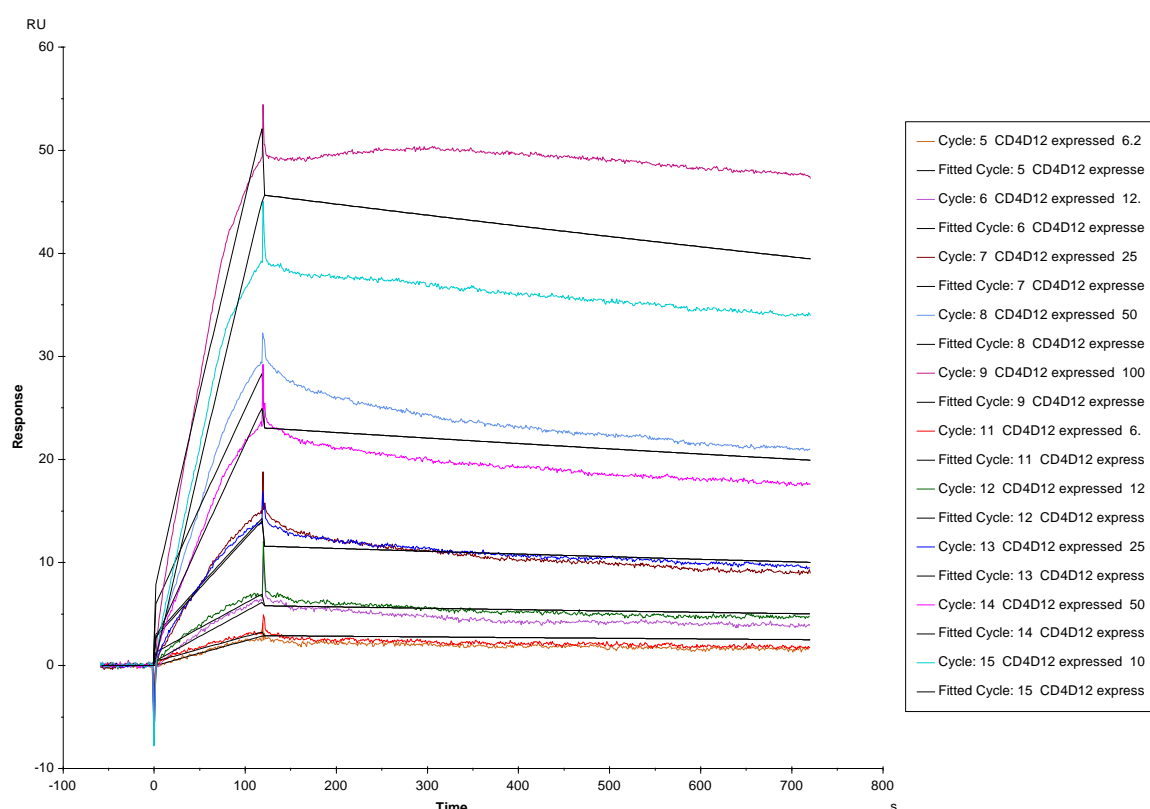


Figure 79 Replicate 1 of interaction of analyte CD4D12ex with immobilised ligand gp120 at 25C. The coloured lines depict the double-referenced sensograms obtained from duplicate injections of 100, 50, 25, 12.5, 6.3 nM analyte CD4D12<sub>ex</sub> across 1000 RU immobilised ligand gp120. The ligand surface was regenerated with a 60 s injection of 10 mM glycine pH 1.75 between each binding cycle. The black lines depict the global fit of the data to a 1:1 Langmuir interaction model, yielding  $k_a = 3.426 \text{ M}^{-1} \cdot \text{s}^{-1}$ ,  $k_d = 2.425 \cdot 10^{-4} \cdot \text{s}^{-1}$ ,  $K_d = 7.08 \cdot 10^{-5} \text{ M}$ .

As can be observed in **Figure 79** the model did not fit the obtained data set in any of the tested concentrations. Moreover, the obtained kinetic constants:  $k_a = 3.426 \text{ M}^{-1} \cdot \text{s}^{-1}$ ,  $k_d = 2.425 \cdot 10^{-4} \cdot \text{s}^{-1}$ ,  $K_d = 7.08 \cdot 10^{-5} \text{ M}$  were not in the range of the published data ( $K_d = 6\text{-}52 \text{ nM}$ ,  $k_a = 5.3 \cdot 10^3\text{-}4.3 \cdot 10^4 \text{ M}^{-1} \cdot \text{s}^{-1}$ , and  $k_d = 2.6 \cdot 10^{-4} \cdot \text{s}^{-1}$ ).<sup>[128][124][130]</sup> To understand the obtained values better, our attention was turned towards the raw data obtained from the SPR experiment. The baseline increased significantly during the experiment and was not regenerated properly (**Figure 80**, top). Moreover, the binding level at higher concentrations (50 and 100 nM, **Figure 80**,

bottom) was not reproducible and decreased significantly, which might have been caused by the analyte built-up on the immobilised gp120. The underlying cause of the obtained results might be the ununiform content of the CD4D12<sub>ex</sub> sample. As it was shown in section 2.1, the expressed CD4D12<sub>ex</sub> was present in two populations: the properly folded and misfolded variant. The misfolded fraction of the protein could be interacting with gp120 differently from the properly folded CD4D12, which in consequence obscured the obtained results.

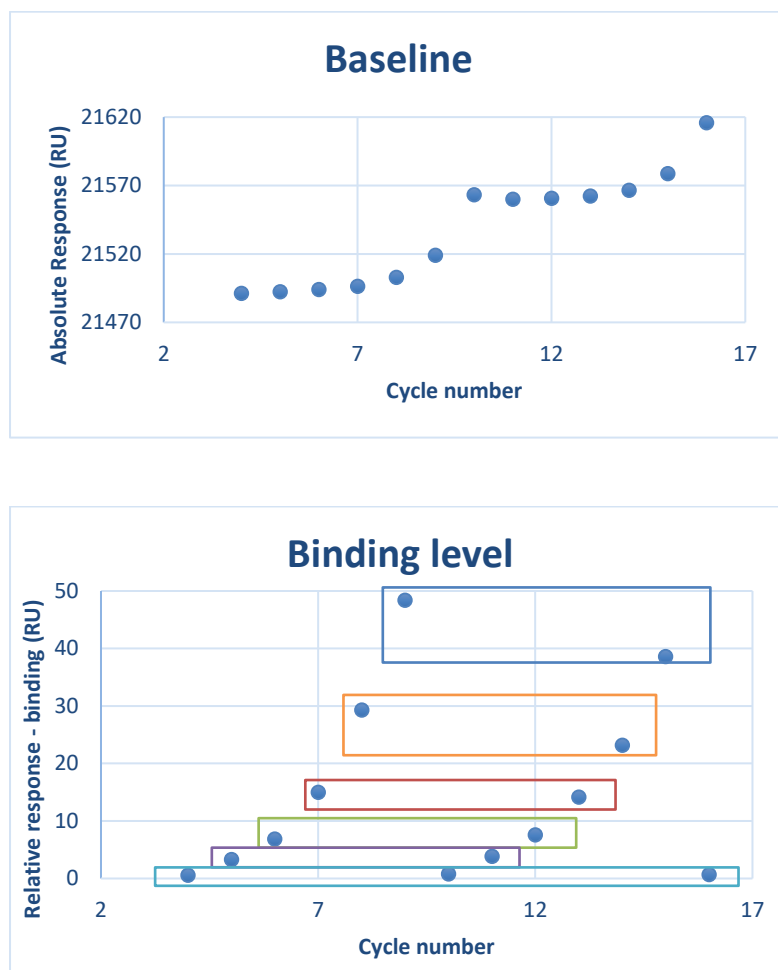


Figure 80 Baseline and binding level in binding CD4D12<sub>ex</sub> to immobilised gp120.12<sub>ex</sub> to immobilised gp120.

Top: Scatter plot of baseline report point values (RU) against consequent cycles. Bottom: Scatter plot of binding level report point values (RU) against consequent cycles. The blue box encircles injection of 100 nM CD4D12<sub>ex</sub>. The orange box encircles injection of 50 nM CD4D12<sub>ex</sub>. The red box encircles injection of 25 nM CD4D12<sub>ex</sub>. The green box encircles injection of 12.5 nM CD4D12<sub>ex</sub>. The purple box encircles injection of 6.25 nM CD4D12<sub>ex</sub>. The cyan box encircles injection of 0 nM CD4D12<sub>ex</sub>.

#### 4.2.1.5 Conclusions

Binding of commercial CD4D12 and expressed CD4D12<sub>ex</sub> with immobilised gp120 was evaluated. In case of commercial CD4D12 the obtained kinetic data were

similar to the published data<sup>[128][124][130]</sup>, however immobilised gp120 was highly unstable and degraded rapidly on the surface of the CM5 chip. gp120 degraded over time and this degradation was accelerated by repeated regeneration cycles. The binding of expressed CD4D12<sub>ex</sub> was obscured by the heterogenous content of the sample, since it probably consisted of both folded and misfolded protein, as was discussed in section 2.1. The obtained kinetic data did not reproduce the published data.<sup>[128][124][130]</sup> Therefore it was decided not to carry out competition assays.

To investigate whether a change in experimental setup would be a remedy to the problems of gp120 instability and loss of its activity, it was decided to immobilise commercial CD4D12, which is described in section 4.2.2. Moreover, immobilisation of CD4D12 instead of gp120 would make gp120 mimics binding experiments to CD4D12 experiments easier to analyse, since this would involve measuring direct binding of the protein mimics to CD4D12.

#### **4.2.2 Towards measuring direct binding to immobilised CD4D12**

In the process of setting up the binding assay of CD4D12 to immobilised gp120 it was found, that the volumes and concentrations of both interacting partners were lower than expected. Therefore, utilisation of gp120 could be used as an analyte. Moreover, if CD4D12 was immobilised on the chip, it would allow preparation of a sample of gp120 immediately before the experiment, which would minimise the chance of gp120 losing activity over time. Immobilisation of CD4D12 also simplified the analysis of the binding capacity of gp120 mimics to CD4D12 easier, since there are less variables that should be taken into account.

In order to test the binding of gp120 mimics to CD4D12, a surface with immobilised CD4D12 had to be prepared. Therefore, the optimal conditions for this immobilisation had to be determined.

##### **4.2.2.1 pH scouting and pre-concentration**

To find optimal coupling conditions for CD4D12 immobilisation, CD4D12 was passed over an unactivated sensor chip under different buffer conditions and the extent of its preconcentration was observed as an increase in response.

Solutions of CD4D12 (concentration of  $10 \mu\text{g}\cdot\text{mL}^{-1}$ ) were prepared in 10 mM acetate buffers of different pH values: 5.5, 5.0, 4.5 and 4.0, and exposed to the unactivated surface of the CM5 chip. After completion of each protein injection, 1 M ethanolamine solution was injected at pH 8.5 to remove the last traces of electrostatically bound ligand. As can be observed in **Figure 81**, the pre-concentration levels of CD4D12 on the chip surface varied at different pH values. At pH 5.5 the obtained pre-concentration level was the lowest (4513.0 RU) and it was observed that the surface was well regenerated. At pH 5.0 the pre-concentration level is high (7146.5 RU) and the baseline is reasonably well regenerated. At pH 4.5 the pre-concentration level is the highest (7573.1 RU), however the surface of the chip does not regenerate well. At pH 4.0 the pre-concentration level is very low (4777.3 RU) and the surface of the chip is also not regenerated. Therefore, the pH values of 4.0-4.5 might have led to protein aggregation or denaturation. The optimal pH for immobilisation was decided to be 5.0.

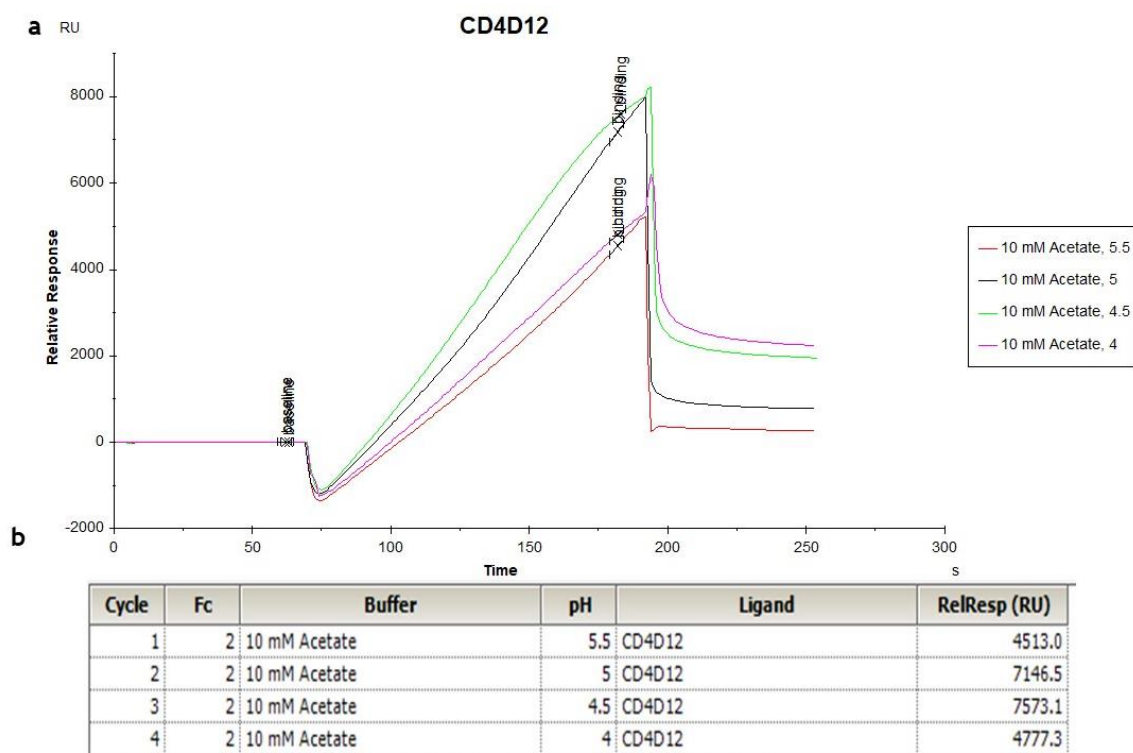


Figure 81 pH scouting and pre-concentration studies of CD4D12.

a) pH scouting and pre-concentration studies of CD4D12. CD4D12 ( $10 \mu\text{g}\cdot\text{mL}^{-1}$ ) in 10 mM acetate buffers at pHs: 5.5, 5.0, 4.5 and 4.0 injections over unactivated CM5 chip over 120 s; b) the table with responses (RU) obtained after each injection of CD4D12.

#### 4.2.2.2 Immobilisation of CD4D12

The approach used for CD4D12 immobilisation was the covalent attachment of the protein to the surface of the sensor chip.

To immobilise CD4D12 on the CM5 chip, firstly, the sensor surface was activated by preparation of the hydroxy succinic ester derivatives of the carboxylic acid moieties on the dextran layer by injection of 0.4 M EDC/0.1 M NHS solution. Activation was followed by injection of a CD4D12 solution at the concentration of  $10\text{ }\mu\text{g}\cdot\text{mL}^{-1}$ . The protein was injected in short pulses until the desired amount of the protein on the surface of the chip was achieved. The last step was the removal of the non-covalently bound ligand and deactivation of the sensor surface by conversion of the remaining active ester groups to amides by injection of 1 M ethanolamine pH 8.5.

A test chip with 1700 RU of CD4D12 immobilised on it was used to find optimal regeneration conditions as well as to investigate binding of gp120 to CD4D12. The chip with 1200 RU of CD4D12 immobilised (**Figure 82**) was used to find optimal experimental conditions for binding of the gp120 mimics, and also in the binding assays of the gp120 mimics.

Flow cell 1 served as a reference surface, therefore it was activated and deactivated similarly to the active surface in flow cell 2, albeit not exposed to CD4D12.

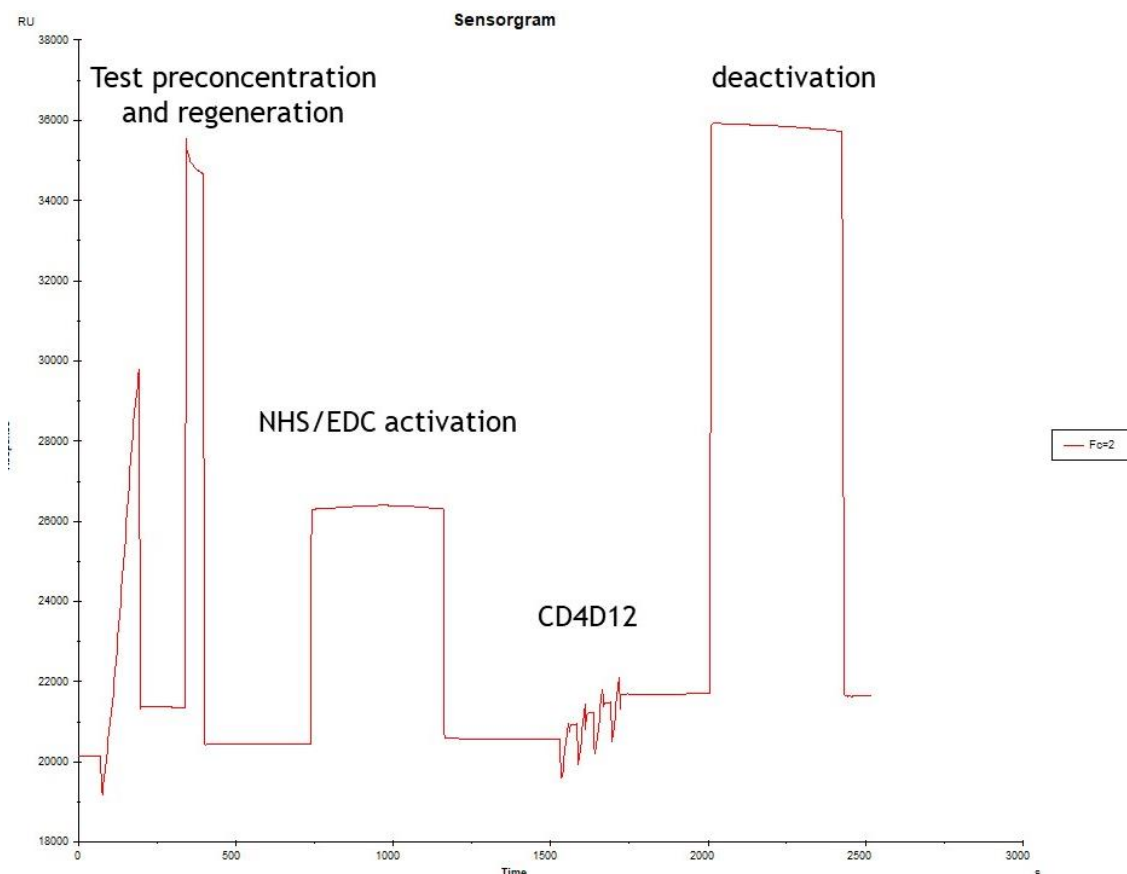


Figure 82 The process of CD4D12 covalent immobilisation on the surface of CM5 chip.

#### 4.2.2.3 Regeneration scouting

Four regeneration buffers were tested, these contained: 10 mM glycine buffers with different pH values: 2.0, 1.75, 1.5 and 10 mM glycine pH 1.75 supplemented with 150 mM NaCl. The regeneration buffer was injected for 30 s. The regeneration was started from testing the mildest regeneration conditions (10 mM glycine pH 2.0). At pH 2.0 an increase in baseline and decrease in the sample response was observed, which suggested that gp120 was not removed completely, and it had accumulated on the sensor surface. At pH 1.75 the baseline was still decreasing while the sample response was steady. The decrease in a baseline might have been caused by removal of the accumulated gp120 from the previous cycle. At pH 1.5 the baseline was still decreasing while the sample response was steady. When two different conditions give similar responses (pH 1.75 and pH 1.5), milder conditions (pH 1.75) should be chosen. At pH 1.75 with added 150 mM NaCl the baseline started increasing again, while the sample response decreased. These results confirmed that 10 mM Glycine at pH 1.75 was an optimal regeneration buffer.

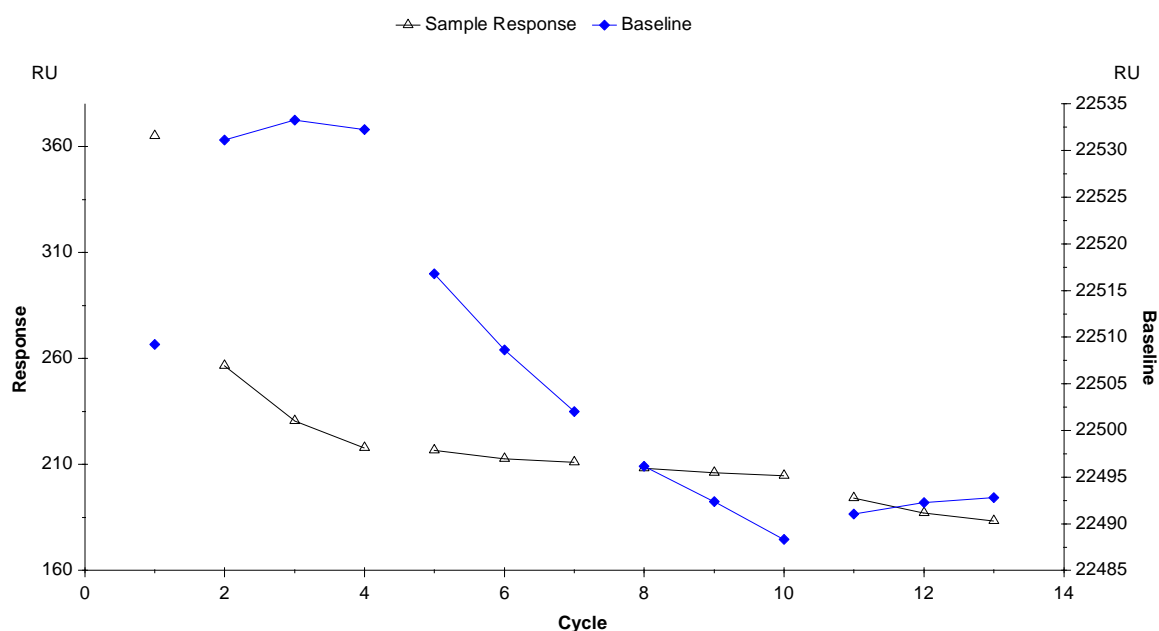


Figure 83 Regeneration scouting of the chip surface with immobilised CD4D12. The ligand CD4D12 was immobilised and gp120 was injected over it at the concentration of 100 nM. The conditions tested from the left: 10 mM glycine pH 2.0, 10 mM glycine pH 1.75, 10 mM, 10 mM glycine pH 1.5, 10 mM glycine pH 1.75 + 150 mM NaCl.

#### 4.2.2.4 Binding of gp120 to the immobilised CD4D12

Binding of gp120 to the immobilised CD4D12 was evaluated to investigate whether CD4D12 is more stable on the sensor surface than gp120, which would then allow to test gp120 mimics. Since in the previous section it was shown, that at concentrations of 250 nM the models did not fit the obtained data sets, the maximum concentration was decreased to 200 nM in the first replicate. Since at this concentration the models did not fit the data either, the highest concentration was decreased further to 100 nM in replicates 2 and 3. The 200 nM concentration from the first replicate was excluded from calculations of the kinetic data.

Serial dilutions (0-100 nM) of gp120 were prepared in 1×HBS-P buffer and injected over 1700 RU immobilised CD4D12 on the CM5 chip. The association time was 120 s, followed by 600 s of dissociation and finally removal by two injections of 10 mM glycine at pH 1.75 for 30 s each. Each concentration of gp120 was injected once, starting with the highest concentration of gp120. A solution of 50 nM gp120 was injected for the second time after all the samples had been evaluated, to determine if the response was at the same level after the entire assay. This experiment was repeated twice. Binding of gp120 to the immobilised CD4D12 was



fitted a 1:1 Langmuir binding model describing the interaction of two molecules in a 1:1 complex.

The sensograms corresponding to the 3 replicates of the experiment are shown in **Figure 84**, **Figure 85** and **Figure 86** and the obtained kinetic data are summarised in **Table 12**.

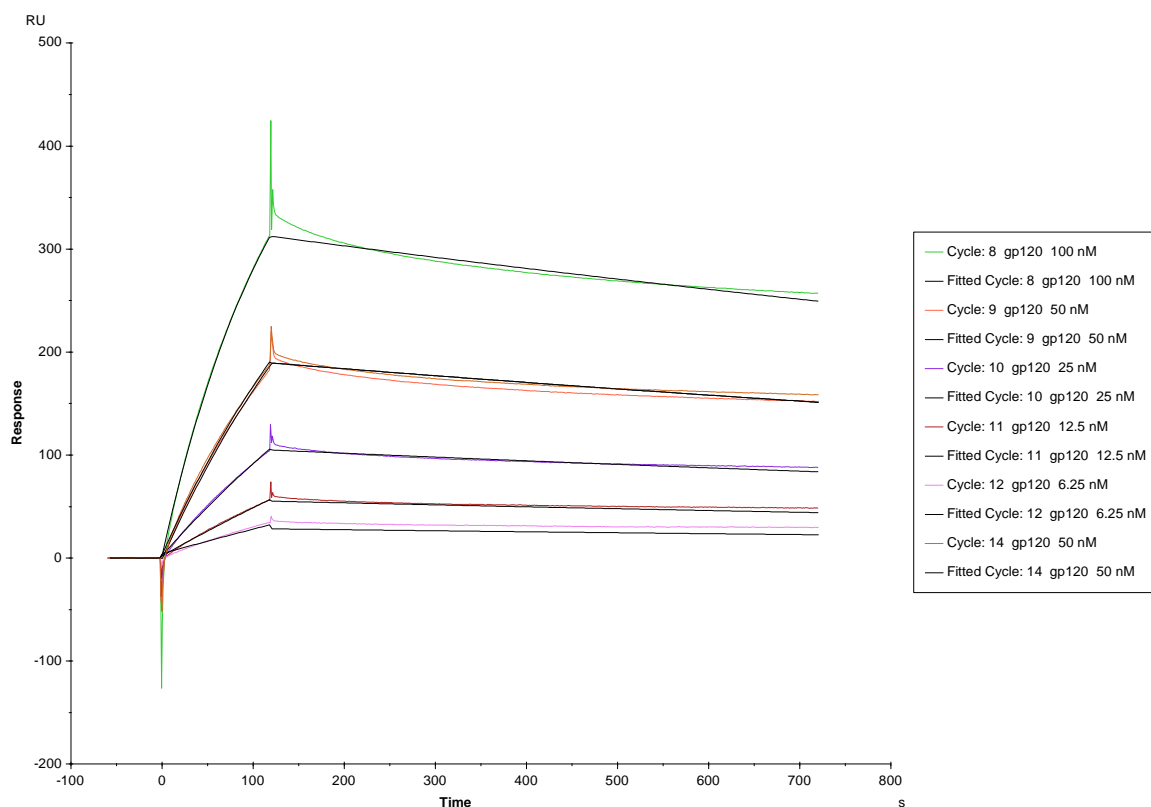


Figure 84 Replicate 1 of the interaction of analyte gp120 with immobilised ligand CD4D12 at 25°C. The coloured lines depict the double-referenced sensograms obtained from injections of 100, 50, 25, 12.5, 6.25 nM analyte gp120 across 1700 RU immobilised ligand CD4D12. Injection of 50 nM gp120 was repeated at the end of the experiment. The ligand surface was regenerated with two 30 s injections of 10 mM glycine pH 1.75 between each binding cycle. The black lines depict the global fit of the data to a 1:1 Langmuir interaction model, yielding  $k_a = 7.274 \cdot 10^4 \text{ M}^{-1} \cdot \text{s}^{-1}$ ,  $k_d = 3.748 \cdot 10^{-4} \cdot \text{s}^{-1}$ ,  $K_d = 5.152 \cdot 10^{-9} \text{ M}$ .

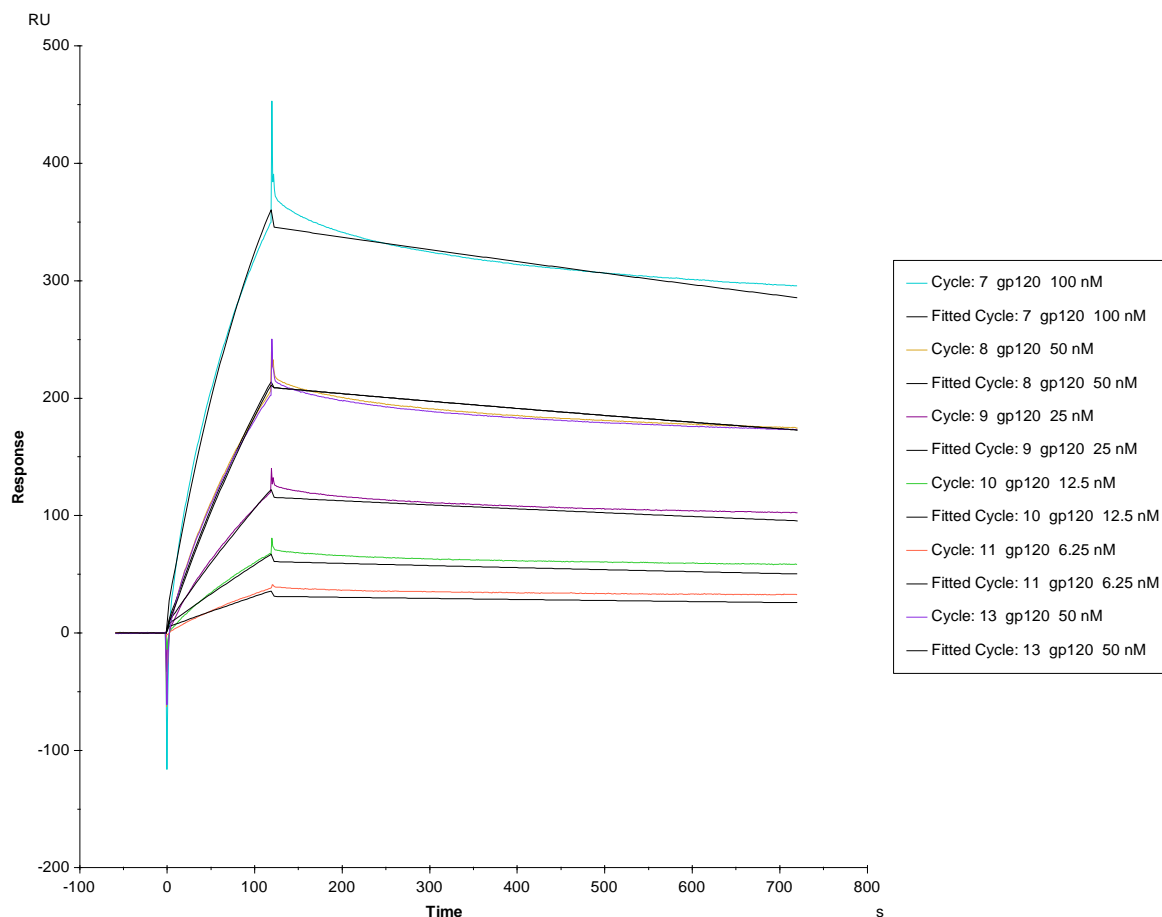


Figure 85 Replicate 2 of the interaction of analyte gp120 with immobilised ligand CD4D12 at 25°C. The coloured lines depict the double-referenced sensograms obtained from injections of 100, 50, 25, 12.5, 6.25 nM analyte gp120 across 1700 RU immobilised ligand CD4D12. Injection of 50 nM gp120 was repeated at the end of the experiment. The ligand surface was regenerated with two 30 s injections of 10 mM glycine pH 1.75 between each binding cycle. The black lines depict the global fit of the data to a 1:1 Langmuir interaction model, yielding  $k_a = 7.332 \cdot 10^4 \text{ M}^{-1} \cdot \text{s}^{-1}$ ,  $k_d = 3.220 \cdot 10^{-4} \cdot \text{s}^{-1}$ ,  $K_d = 4.392 \cdot 10^{-9} \text{ M}$ .

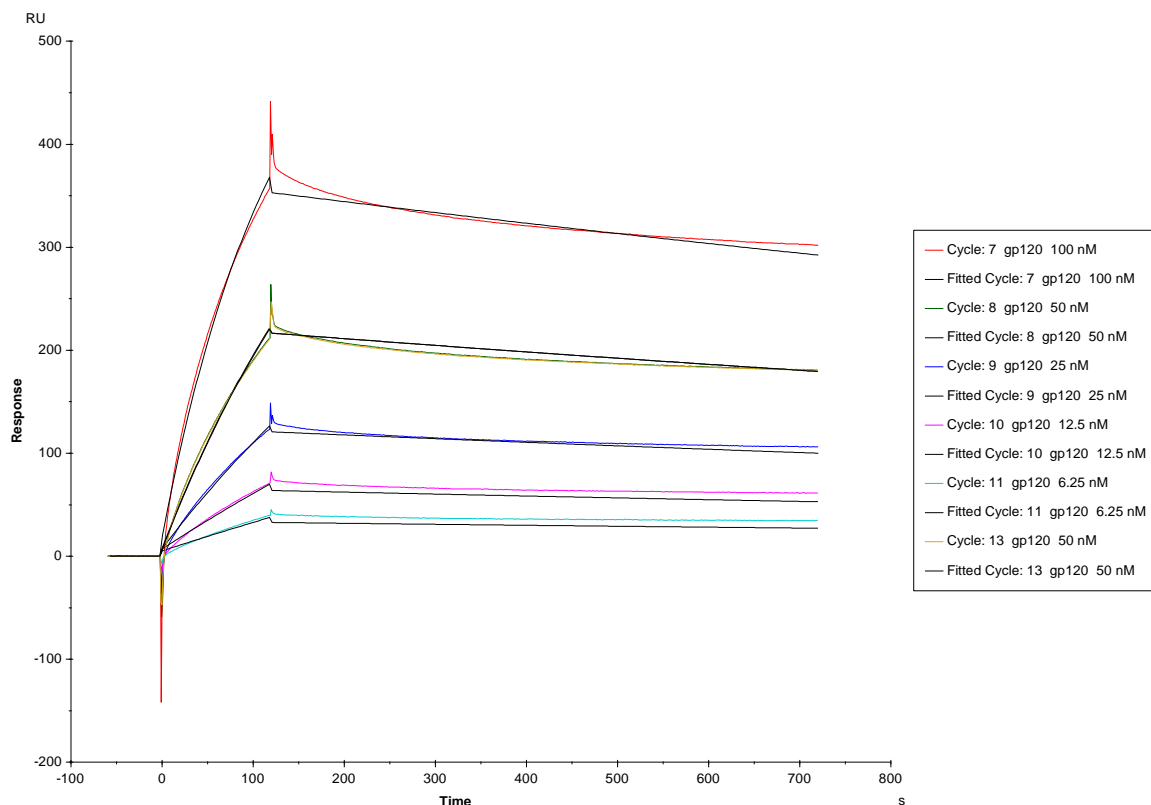


Figure 86 Replicate 3 of the interaction of analyte gp120 with immobilised ligand CD4D12 at 25°C. The coloured lines depict the double-referenced sensograms obtained from injections of 100, 50, 25, 12.5, 6.25 nM analyte gp120 across 1700 RU immobilised ligand CD4D12. Injection of 50 nM gp120 was repeated at the end of the experiment. The ligand surface was regenerated with two 30 s injections of 10 mM glycine pH 1.75 between each binding cycle. The black lines depict the global fit of the data to a 1:1 Langmuir interaction model, yielding  $k_a = 7.775 \cdot 10^4 \text{ M}^{-1} \cdot \text{s}^{-1}$ ,  $k_d = 3.141 \cdot 10^{-4} \text{ s}^{-1}$ ,  $K_d = 4.039 \cdot 10^{-9} \text{ M}$ .

Table 12 Kinetic data obtained from the three experiments of binding gp120 to immobilised CD4D12.  $k_a$  – association rate,  $k_d$  – dissociation rate,  $K_d$  – equilibrium dissociation, constant,  $R_{\text{max}}$  – analyte binding capacity.

Replicate	$k_a \text{ (M}^{-1} \cdot \text{s}^{-1})$	$k_d \text{ (s}^{-1})$	$K_d \text{ (M)}$	$R_{\text{max}} \text{ (RU)}$
1	$7.274 \cdot 10^4$	$3.748 \cdot 10^{-4}$	$5.152 \cdot 10^{-9}$	546.9
2	$7.332 \cdot 10^4$	$3.220 \cdot 10^{-4}$	$4.392 \cdot 10^{-9}$	606.3
3	$7.775 \cdot 10^4$	$3.141 \cdot 10^{-4}$	$4.039 \cdot 10^{-9}$	591.4
Mean	$7.46 \cdot 10^4$	$3.37 \cdot 10^{-4}$	$4.53 \cdot 10^{-9}$	-
Standard deviation	3.7%	9.8%	12.6%	-

As can be observed from **Table 12** the obtained kinetic values were consistent with the previously published kinetic data for gp120-CD4D12 interaction ( $K_d = 6-52$  nM,  $k_a = 5.3 \cdot 10^3 - 4.3 \cdot 10^4$   $M^{-1} \cdot s^{-1}$ , and  $k_d = 2.6 \cdot 10^{-4} \cdot s^{-1}$ ).<sup>[128][124][130]</sup>  $K_d$ 's of 4.5 nM,  $k_a = 7.775 \cdot 10^4$   $M^{-1} \cdot s^{-1}$  and  $k_d = 3.37 \cdot 10^{-4} \cdot s^{-1}$  (mean values) were obtained. This was also a significant improvement over the previous experimental setup with gp120 immobilised on the sensor chip. These results also indicated, that in contrast to immobilised gp120, immobilised CD4D12 was more stable on the chip and could be reused in several assays successfully.

With the successfully obtained positive control of gp120-CD4D12 interaction and the found good experimental setup, which yielded reproducible kinetic data it was now possible to perform experiments with gp120 mimics.

#### 4.2.2.5 Binding of gp120 mimics to the immobilised CD4D12

The binding of the synthesized constructs **22**, **26** and **30** (**Figure 87**) to the immobilised CD4D12 was tested.

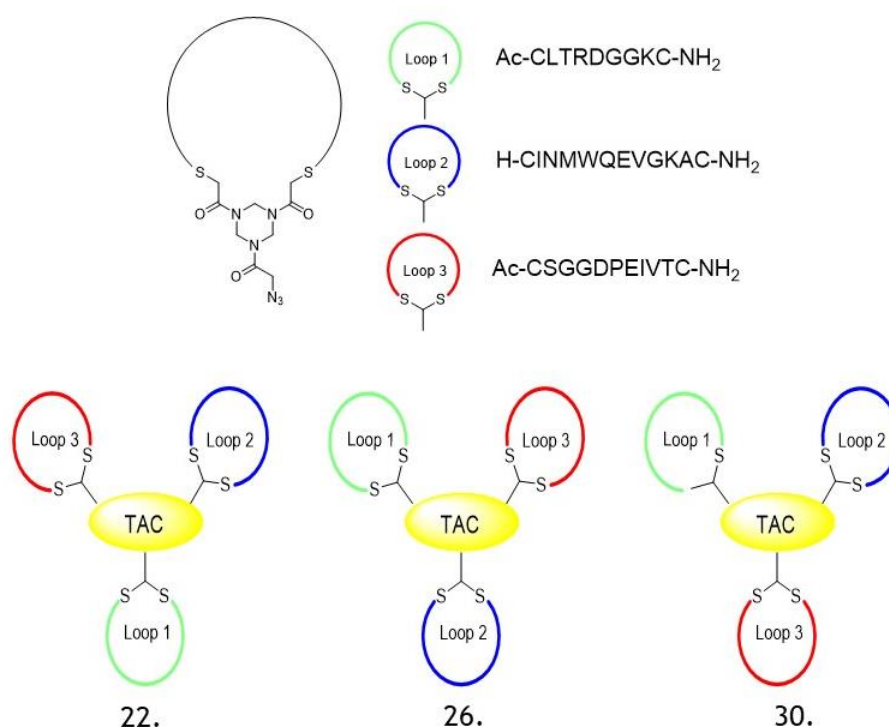


Figure 87 Upper: Schematic representation of loops 1, 2 and 3 structures; lower: schematic representation of the synthesized gp120 discontinuous mimics **22**, **26** and **30** structures.

Firstly, the experimental conditions for evaluating the gp120 constructs were the same as for binding of gp120 to the immobilised CD4D12, and construct **30** was used first since it was obtained in the highest amount. However, it was found that

by using the same conditions, negative unprocessed sensograms were obtained both in the reference flow cell and in the flow cell with immobilised ligand, which did not allow to calculate the kinetic data (**Figure 88**, left and right).

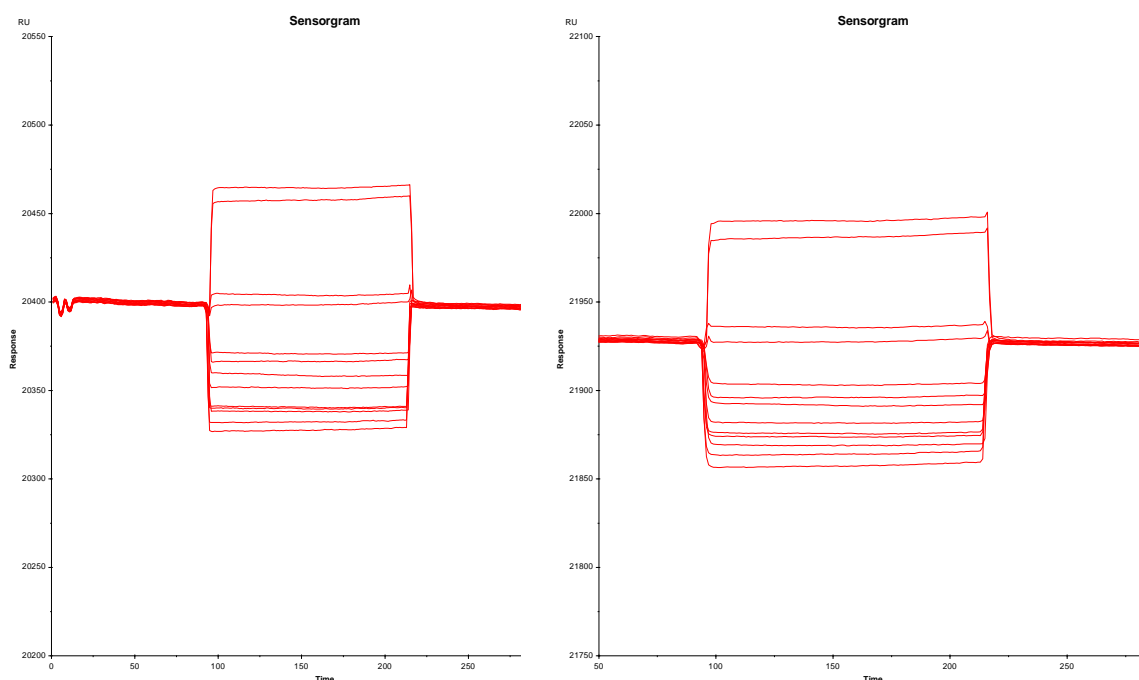


Figure 88 Unprocessed sensorgrams of construct **30** binding to: left: the blank sensor surface; right: the immobilised ligand CD4D12.

One of the possible explanations would be binding of the mimic **30** to the blank sensor surface instead of to the ligand, which would result in the negative sensorgram. However, no binding to the reference surface was observed (**Figure 89**, top). Moreover, the binding level throughout the experiment was not consistent (**Figure 89**, bottom).

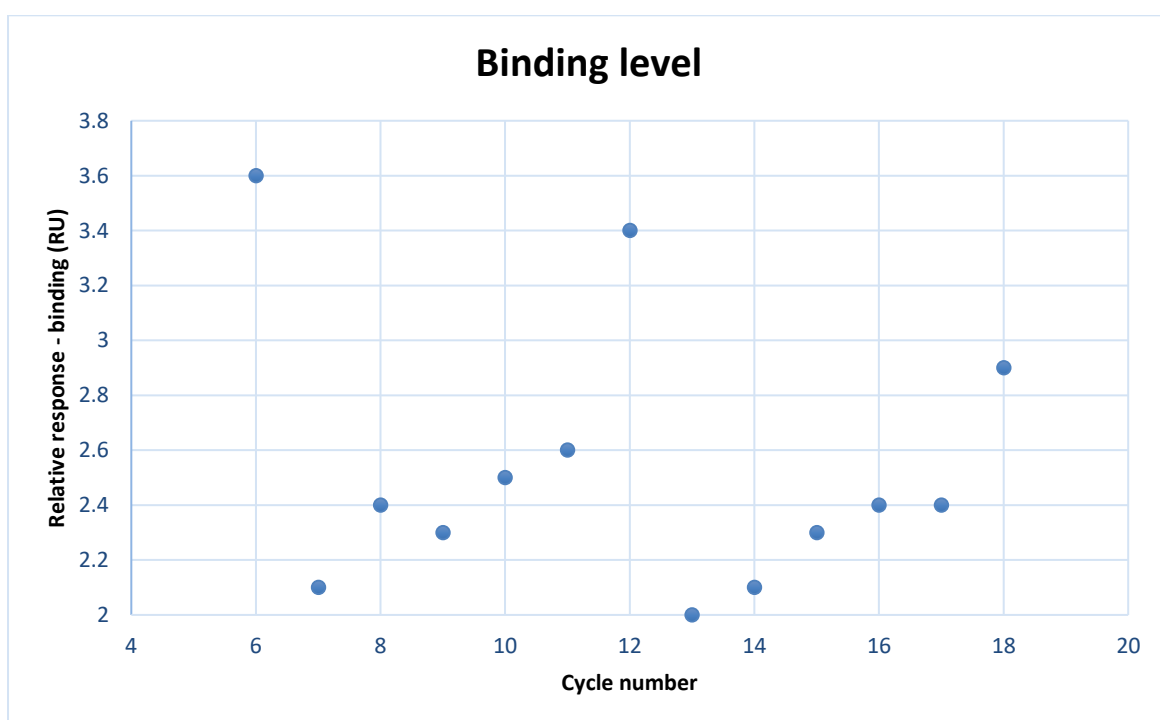
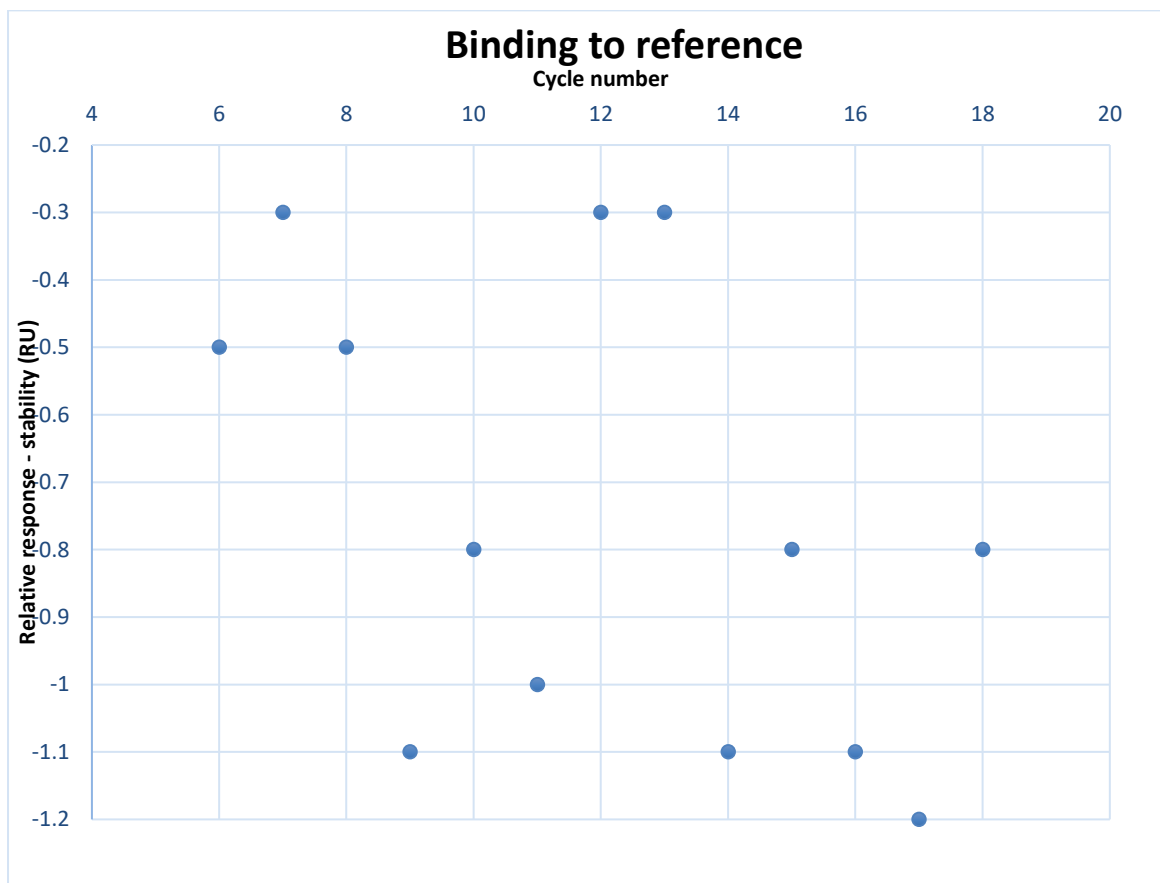


Figure 89 Top: Binding to the reference channel (RU) in each cycle, bottom: binding level responses (RU) in each cycle.

Since the most obvious cause (binding to the reference surface) was eliminated, it was decided to change the running buffer from HBS-P to PBS-P. The Biacore™

Assay Handbook<sup>[161]</sup> suggested that HBS-P buffer is preferred in protein assays, while PBS-P is preferred in small molecule assays, as HEPES could bind the ligand and interfere with detection of low molecular weight compounds. Change of the buffer was a major improvement, and all the constructs could be tested.

gp120 protein mimics are significantly smaller when compared to gp120 (5 kDa and 120 kDa, respectively). SPR monitors the change in mass on the surface of the chip, therefore the larger the molecular weight of the analyte, the easier it is to obtain a response. Therefore, it was decided to increase the highest concentration of tested constructs from 100 nM to 100  $\mu$ M. Moreover, recently published  $IC_{50}$  values for similar constructs were in the range of 41.3-57.3  $\mu$ M.<sup>[88]</sup> The highest concentration that could be obtained to repeat each experiment and test each concentration in duplicate was 100  $\mu$ M, which was due to the limited amount of the synthesized constructs.

Serial dilutions of gp120 mimics **22**, **26** and **30** were prepared (0-100  $\mu$ M) in 1 $\times$ PBS-P buffer and injected over 1200 RU immobilised CD4D12 on the CM5 chip. The association time was 120 s, followed by 600 s of dissociation and finally removed by two injections of 10 mM glycine at pH 1.75 for 30 s each. Each experiment was performed in a duplicate. Since CD4D12 proved to be stable on the sensor surface, each concentration of gp120 mimic was injected twice, starting from the lowest concentration to the highest. Binding of the gp120 mimics to the immobilised CD4D12 was fitted a 1:1 Langmuir binding model describing the interaction of two molecules in a 1:1 complex.

Firstly, construct **22** was tested for binding to CD4D12. Binding of construct **22** was very weak even at the highest (100  $\mu$ M) concentration. The selected obtained sensogram is shown in **Figure 90** (the sensogram of the duplicated experiment is attached in the Appendix **Figure 168**). The obtained kinetic constants were as follow:  $k_a = 39.40 \text{ M}^{-1} \cdot \text{s}^{-1}$ ,  $k_d = 4.203 \cdot 10^{-4} \cdot \text{s}^{-1}$ ,  $K_d = 1.067 \cdot 10^{-5} \text{ M}$ . However, the reported kinetic constant  $k_a$  was outside the limits that could be measured by the instrument (detection limit of Biacore™ X100 for  $k_a$ :  $10^3$ - $10^7 \text{ M}^{-1} \text{ s}^{-1}$  and for  $k_d$ :  $10^{-5}$ - $0.1 \text{ s}^{-1}$  as reported in the Biacore™ X100 handbook<sup>[162]</sup>), therefore this value unfortunately could not be considered a valid result.

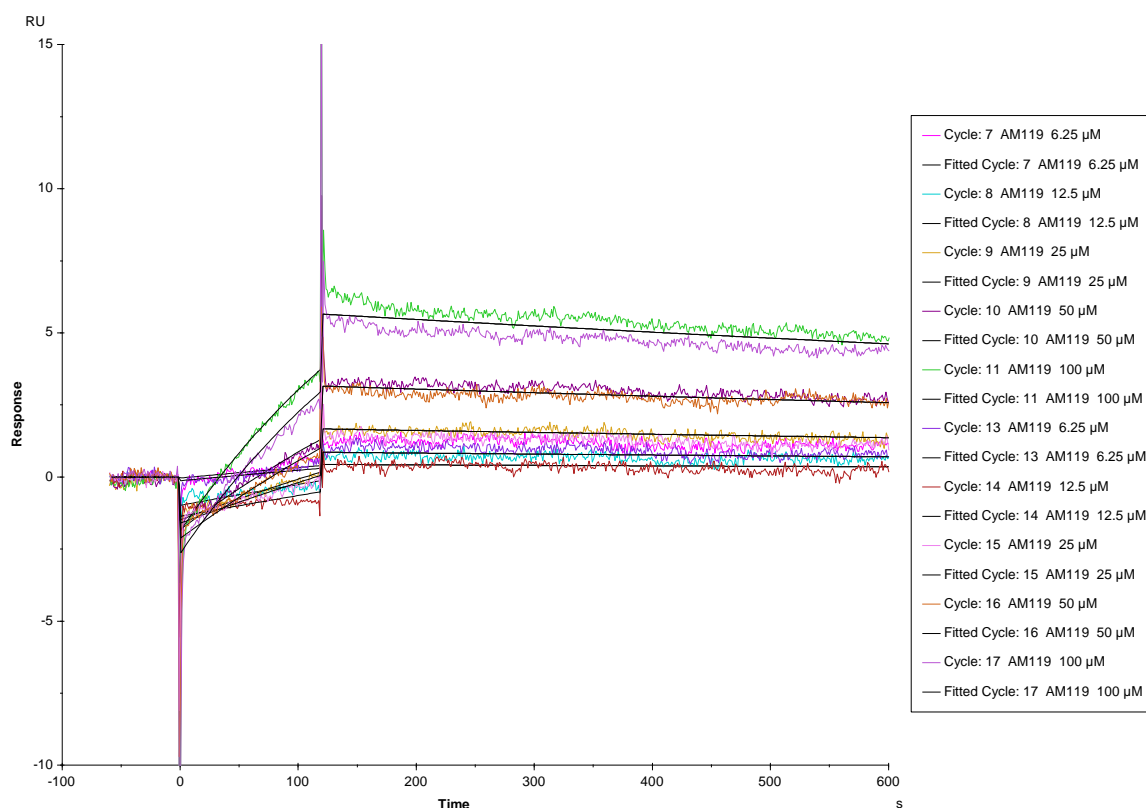


Figure 90 Replicate 1 of performed experiment to measure binding of mimic **22** to the immobilised CD4D12 at 25C.

The coloured lines depict the double-referenced sensograms obtained from duplicate injections of 100, 50, 25, 12.5, 6.25  $\mu\text{M}$  analyte **22** across 1200 RU immobilised ligand CD4D12. The ligand surface was regenerated with two 30 s injections of 10 mM glycine pH 1.75 between each binding cycle. The black lines depict the global fit of the data to a 1:1 Langmuir interaction model, yielding  $k_a = 30.40 \text{ M}^{-1}\cdot\text{s}^{-1}$ ,  $k_d = 4.203\cdot 10^{-4}\cdot\text{s}^{-1}$ ,  $K_d = 1.067\cdot 10^{-5} \text{ M}$ .

Binding of the second tested construct **26** was very weakly detectable even at the highest 100  $\mu\text{M}$  concentration. The selected obtained sensogram is shown in Figure **91** (the sensogram of the duplicated experiment is attached in the Appendix **Figure 169**). The kinetic constants could not be uniquely determined for this construct.



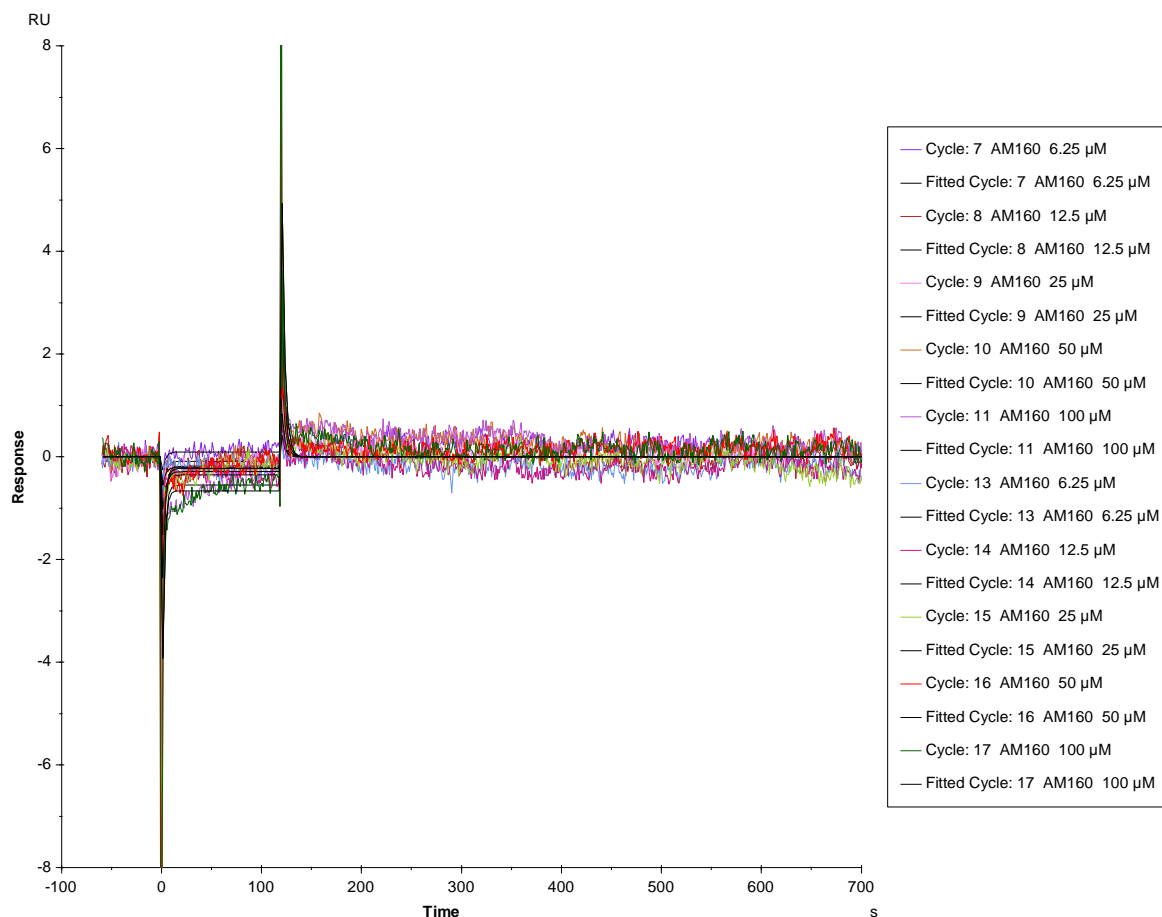


Figure 91 Replicate 1 of performed experiment to measure binding of mimic **26** to the immobilised CD4D12 at 25C.

The coloured lines depict the double-referenced sensograms obtained from duplicate injections of 100, 50, 25, 12.5, 6.25  $\mu$ M analyte **26** across 1200 RU immobilised ligand CD4D12. The ligand surface was regenerated with two 30 s injections of 10 mM glycine pH 1.75 between each binding cycle. The black lines depict the global fit of the data to a 1:1 Langmuir interaction model.

Binding of the last tested construct **30** was very weak even at the highest 100  $\mu$ M concentration. This construct showed intermediate potency between constructs **22** and **26**. The selected obtained sensogram is shown in **Figure 92** (the sensogram of the duplicated experiment is attached in the Appendix **Figure 170**). The obtained kinetic constants were as follow:  $k_a = 4.603 \cdot 10^5 \text{ M}^{-1} \cdot \text{s}^{-1}$ ,  $k_d = 5.197 \cdot 10^{-4} \cdot \text{s}^{-1}$ ,  $K_d = 1.129 \cdot 10^{-9} \text{ M}$ . However, care has to be taken when reviewing the obtained kinetic constant values.

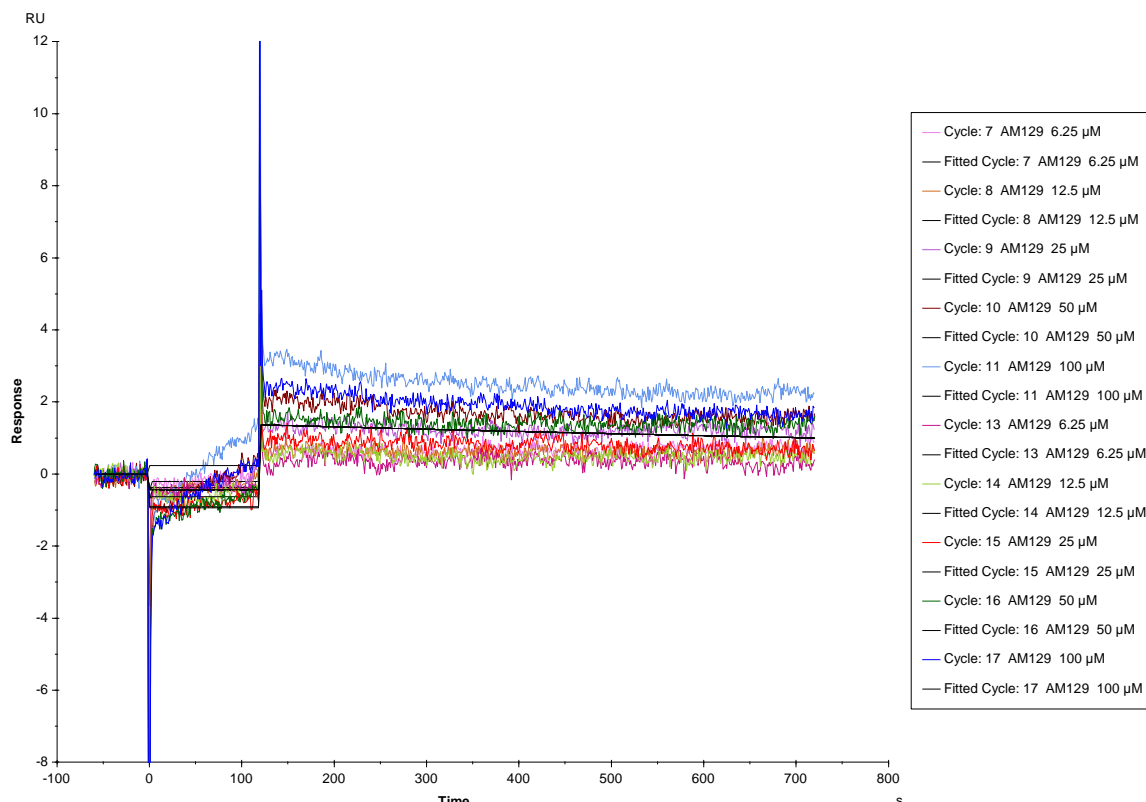


Figure 92 Replicate 1 of performed experiment to measure binding of mimic **30** to the immobilised CD4D12 at 25C.

The coloured lines depict the double-referenced sensograms obtained from duplicate injections of 100, 50, 25, 12.5, 6.25  $\mu\text{M}$  analyte **30** across 1200 RU immobilised ligand CD4D12. The ligand surface was regenerated with two 30 s injections of 10 mM glycine pH 1.75 between each binding cycle. The black lines depict the global fit of the data to a 1:1 Langmuir interaction model, yielding  $k_a = 4.603 \cdot 10^5 \text{ M}^{-1} \cdot \text{s}^{-1}$ ,  $k_d = 5.197 \cdot 10^{-4} \cdot \text{s}^{-1}$ ,  $K_d = 1.129 \cdot 10^{-9} \text{ M}$ .

#### 4.2.2.6 Binding of independent loops to CD4D12

The binding of synthesized loops 1, 2 and 3 (**Figure 87**) to the immobilised CD4D12 was also evaluated.

Loops mimicking the discontinuous epitope of gp120 are similarly to constructs **22**, **26** and **30** significantly smaller in terms of molecular weight when compared to the protein gp120 (1 kDa and 120 kDa, respectively). Therefore, the tested range of concentrations was the same as for the constructs, 0-100  $\mu\text{M}$ .

Loops 1, 2 and 3 were prepared in serial dilutions (0-100  $\mu\text{M}$ ) in 1×PBS-P buffer and injected independently over 1200 RU immobilised CD4D12 on the CM5 chip. The association time was 120 s, followed by 600 s of dissociation and finally removed by two injections of 10 mM glycine at pH 1.75 for 30 s each. Each experiment was performed in duplicate and each concentration of the loop was injected twice, starting from the lowest to the highest. Binding of the loops to the

immobilised CD4D12 was fitted a 1:1 Langmuir binding model describing the interaction of two molecules in a 1:1 complex.

Firstly, loop 1 was tested for binding to CD4D12. Binding of loop 1 was not detectable even at the highest 100  $\mu\text{M}$  concentration. The selected obtained sensogram is shown on **Figure 93** (the sensogram of the duplicated experiment is attached in the Appendix **Figure 171**). The kinetic constants could not be uniquely determined for this loop.

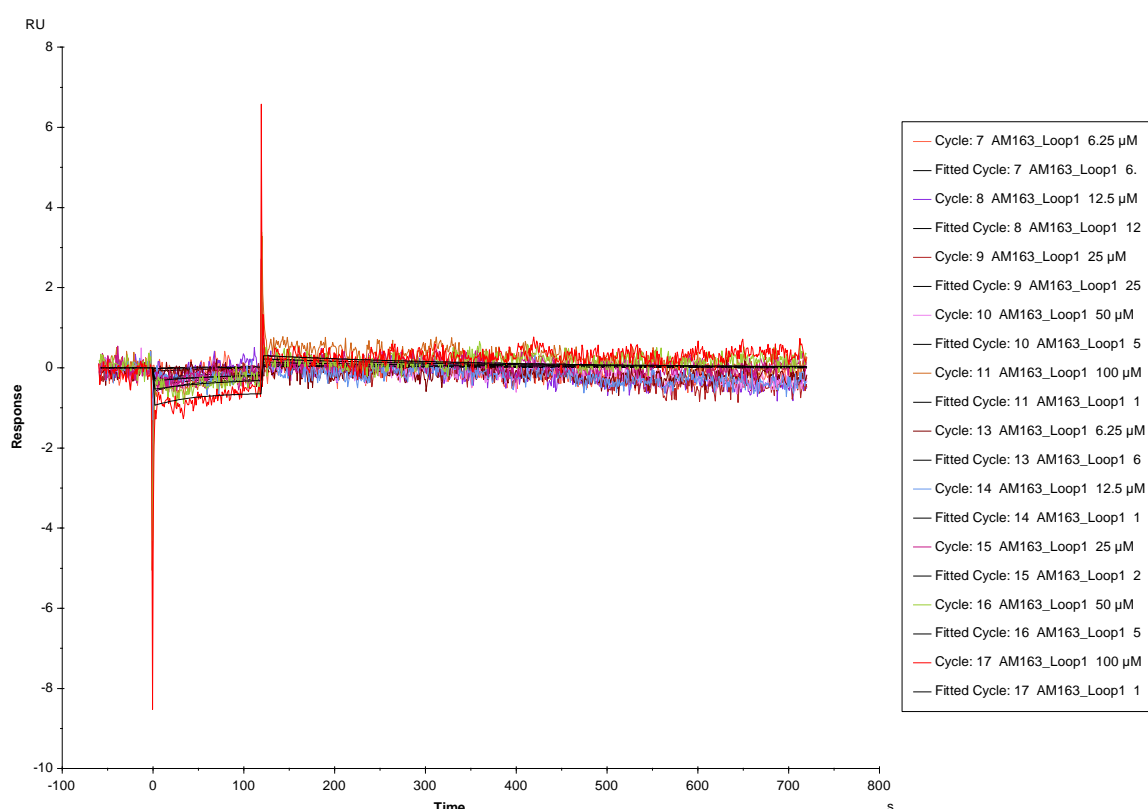


Figure 93 Replicate 1 of performed experiment to measure binding of loop 1 to the immobilised CD4D12 at 25C.

The coloured lines depict the double-referenced sensograms obtained from duplicate injections of 100, 50, 25, 12.5, 6.25  $\mu\text{M}$  analyte loop 1 across 1200 RU immobilised ligand CD4D12. The ligand surface was regenerated with two 30 s injections of 10 mM glycine pH 1.75 between each binding cycle. The black lines depict the global fit of the data to a 1:1 Langmuir interaction model.

Next, loop 2 was tested for binding to the immobilised CD4D12. Interestingly, this loop showed the highest potency out of all the tested constructs or independent loops. The selected obtained sensogram is shown on **Figure 94** (the sensogram of the duplicated experiment is attached in the Appendix **Figure 172**). The obtained kinetic constants were as follow:  $k_a = 226.9 \text{ M}^{-1} \cdot \text{s}^{-1}$ ,  $k_d = 1.229 \cdot 10^{-3} \cdot \text{s}^{-1}$ ,  $K_d = 5.418 \cdot 10^{-6} \text{ M}$ . However, the reported kinetic constant  $k_a$  was outside the limits that could be measured by the instrument, therefore this value unfortunately could not be considered a valid result.

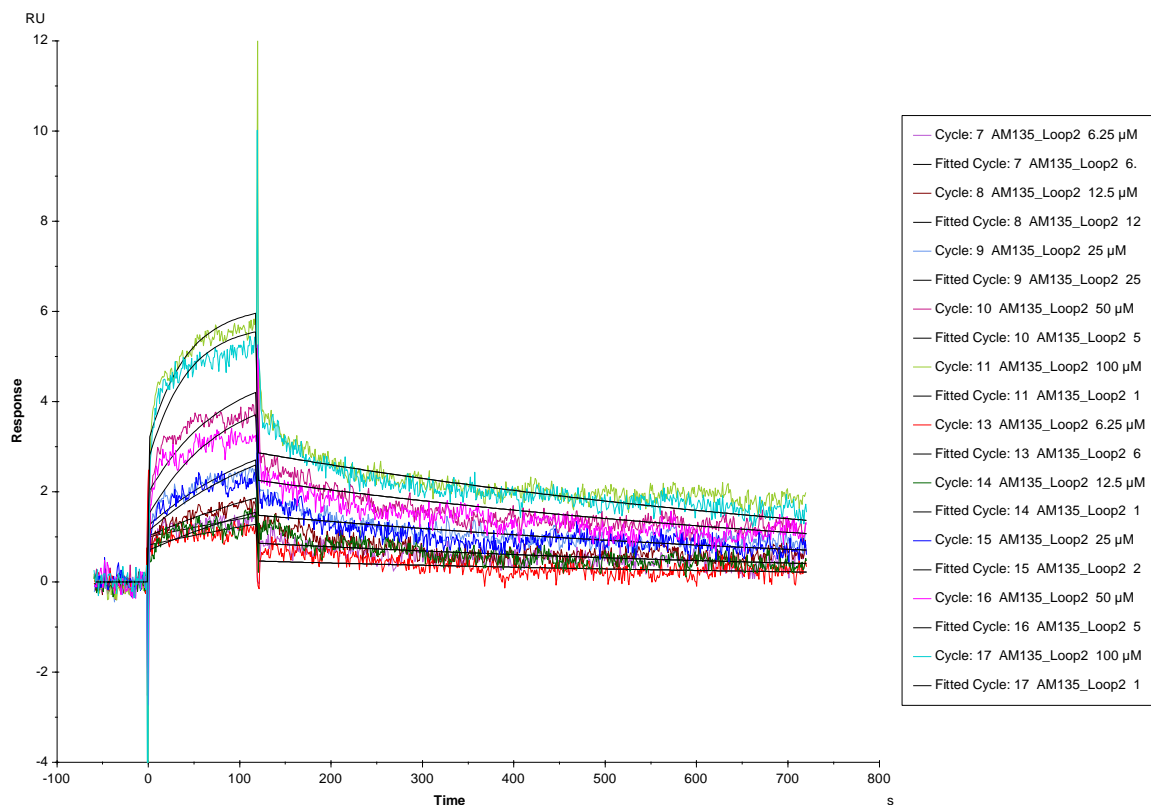


Figure 94 Replicate 1 of performed experiment to measure binding of loop 2 to the immobilised CD4D12 at 25C.

The coloured lines depict the double-referenced sensograms obtained from duplicate injections of 100, 50, 25, 12.5, 6.25  $\mu\text{M}$  analyte loop 2 across 1200 RU immobilised ligand CD4D12. The ligand surface was regenerated with two 30 s injections of 10 mM glycine pH 1.75 between each binding cycle. The black lines depict the global fit of the data to a 1:1 Langmuir interaction model, yielding  $k_a = 226.9 \text{ M}^{-1}\cdot\text{s}^{-1}$ ,  $k_d = 1.229\cdot 10^{-3}\cdot\text{s}^{-1}$ ,  $K_d = 5.418\cdot 10^{-6} \text{ M}$ .

Lastly, loop 3 was tested for binding to CD4D12. Binding of loop 3 was not detectable even at the highest 100  $\mu\text{M}$  concentration. The selected obtained sensogram is shown on **Figure 95** (the sensogram of the duplicated experiment is attached in the Appendix **Figure 173**). The kinetic constants could also not be uniquely determined for this loop.

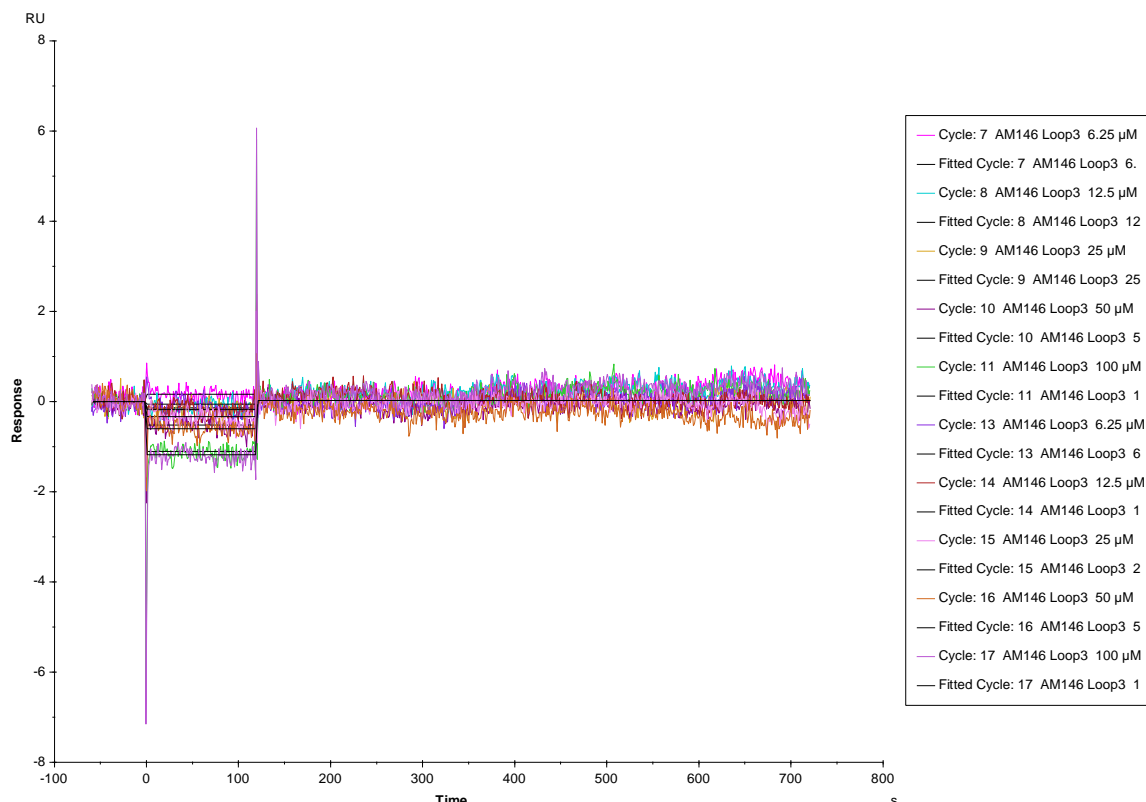


Figure 95 Replicate 1 of performed experiment to measure binding of loop 3 to the immobilised CD4D12 at 25C.

The coloured lines depict the double-referenced sensograms obtained from duplicate injections of 100, 50, 25, 12.5, 6.25  $\mu$ M analyte loop 3 across 1200 RU immobilised ligand CD4D12. The ligand surface was regenerated with two 30 s injections of 10 mM glycine pH 1.75 between each binding cycle. The black lines depict the global fit of the data to a 1:1 Langmuir interaction model.

## 4.3 Conclusions

Binding of gp120 to immobilised CD4D12 as a positive control was successfully optimised by immobilisation of the latter. The obtained kinetic constants of binding of the two proteins were consistent with the published kinetic data of gp120-CD4D12 interaction.<sup>[128][124][130]</sup> Therefore, a reliable method for investigation of gp120 mimics to CD4D12 was developed. Unfortunately, all tested compounds proved to be inactive (construct **26**) or very weakly active (constructs **22** and **30**). This was followed by testing of the binding of the independent loops to CD4D12. Loops **1** and **3** were inactive, while loop **2**, interestingly, showed the highest potency out of the all tested loops and constructs. However, the obtained results have to be reviewed with some care, as the obtained kinetic constant  $k_a$  was outside of the limit of the instrument and the obtained responses were very low.

The difference between the constructs synthesized within this project and recently published constructs<sup>[88]</sup> lies in the used cyclisation hinge. As initial

binding studies showed activity of compounds based on N<sub>3</sub>-DBMB hinge (**Figure 96**) it was decided to prepare compounds with small structural alterations (change of a cyclisation hinge) that would improve their chemical properties. In order to improve the solubility of the cyclic peptides the N<sub>3</sub>-TADB cyclisation hinge was used instead of N<sub>3</sub>-DBMB.

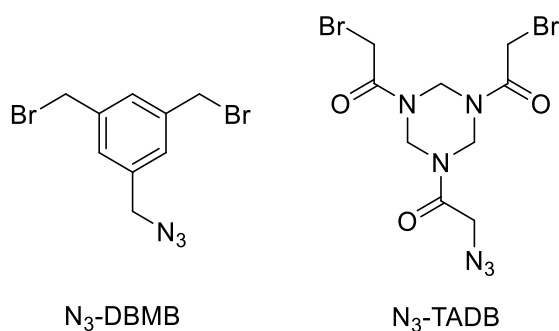


Figure 96 Structures of the cyclisation linkers. Left: N<sub>3</sub>-DBMB; right: (polar hinge) N<sub>3</sub>-TADB.

The polar hinge N<sub>3</sub>-TADB increased significantly the polarity of the synthesized peptide loops, however similarly to the recently published results by van de Langemheen et al.<sup>[155]</sup> it decreased the potency of the obtained products. This trend can be explained by an increased flexibility of the N<sub>3</sub>-TADB hinge when compared to N<sub>3</sub>-DBMB. The N<sub>3</sub>-DBMB hinge obtains planar conformation and it has one methylene unit between the benzene ring and the attached peptide. The N<sub>3</sub>-TADB hinge on the other hand, is present in a chair-like conformation and has one methylene unit more, which introduces more flexibility into the system. Taken together with the flipping cyclophane ring of the TAC scaffold, it might be concluded that the entire molecule is not constrained enough and too flexible to fit into the gp120 binding site of CD4D12. These results show, that computer modelling supplemented with molecular docking to predict preferred orientation of gp120 mimics towards CD4D12 should be approached. Such studies would be helpful, as they would allow to study in greater depth how the individual peptide loops or constructs fit into the binding pocket of CD4D12. In this way it would be possible to assess the correct distance between the TAC scaffold and the loops, the size of the loops (whether they should be bigger or smaller), as well as their rigidity and geometry. Moreover, as could be observed in the experiments of binding of independent loops, apart from loop 2, none of the loops bound CD4D12. Since loop 2 was the most hydrophobic out of the three loops, therefore it was the most capable of creating hydrophobic interactions with CD4D12. Furthermore, since Trp<sup>427</sup> of gp120 is one of the residues contributing in majority to the binding,

it can be assumed that this residue adopts correct geometry and is responsible for binding of loop 2 to CD4D12. The other two residues of gp120 contributing to the binding of CD4D12 are Asp<sup>368</sup> and Glu<sup>370</sup>, which are present in the sequence of loop 3. However, this loop showed no binding towards CD4D12, which allows to draw conclusion, that its design is not optimal. These results do not give insight whether the recognition is sequence specific, therefore it would be beneficial in the future to scramble the peptide sequence and test its binding properties.

Previous studies done in the Liskamp group based on the same concept, however with slightly altered scaffold or cyclisation linker, showed that binding of the loops or constructs to CD4D12 is highly dependent on the used methodology. In the course of this thesis, loop 2 showed the highest binding potential which is in contrast to the previously described results.<sup>[142][144]</sup> The results published by Werkhoven et al.<sup>[88]</sup> with the gp120 mimics that resemble the most closely constructs synthesised and tested in the course of this thesis had the IC<sub>50</sub> values of 41.3 - 57.3  $\mu$ M, with the compound having loop 2 in the “middle” position showing the highest inhibitory potential. Its structural counterpart however, compound **26**, showed no binding potential. Compound **22**, which out of the three constructs showed the best binding properties, had loop 1 in the “middle” position, and loop 2 in the “right” position. Compound **30**, which showed very weak binding, had loop 3 in the “middle” position and, similarly to construct **22**, loop 2 was in the “right” position of the TAC scaffold. Since left and right positions are interchangeable as it was discussed chapter 3.2.3, it can be concluded that loop 2 positioned in either of these positions is responsible for binding to CD4D12. However, if the loop and scaffold size were optimised, it could be possible that binding to CD4D12 of the new constructs would have increased. Results published by Werkhoven et al.<sup>[88]</sup> suggest that cyclic compounds offer small benefits over the use of linear peptides when it comes to their activity, however cyclic compounds have better proteolytic stability in serum, which makes them more attractive than their linear counterparts.

A synergistic effect of assembled loops on the TAC scaffold was expected, however the obtained results indicated, that the attachment of the peptides corresponding to the discontinuous epitopes of gp120 is not enough to obtain an active gp120 mimic. A possible remedy to the flexibility of the TAC scaffold would be utilisation

of a highly pre-organised CTV scaffold.<sup>[108]</sup> Moreover, it would be interesting to design and test other cyclisation hinges, which would allow to solubilise the cyclic peptides without decreasing their potency.

One of the possible explanations of the lack of binding of loops and constructs to CD4D12, was the fact that the protein was immobilised on the chip by covalent binding of the amines at its N-terminus. According to the literature,<sup>[124][128][130][158][159]</sup> this is the most often utilised method, however, gp120 binding-site is located in the proximity of the N-terminus. What is also possible, is that CD4D12 was immobilised heterologously, binding to the chip with other amine residues than the ones present in the N-terminus. To ensure that gp120, gp120 mimics or the loops bind without obstacles to CD4D12, future wise it could be beneficial to immobilise CD4D12 with its C-terminus. Still more, it would be advantageous to attach a linker via the carboxylic moiety of the TAC scaffold that would allow its immobilisation on the chip. In this way, the SPR experiments would benefit, as a “heavy” CD4D12 protein (21 kDa) would be binding to the ligand, which should give better signal response than binding of a “light” analyte - gp120 mimics (4.7 kDa) to CD4D12. Another possibility would be to use MST rather than SPR, as it is highly sensitive technique which requires only very small amounts of the samples.



## 5 Conclusions and future work

Protein-protein interactions play a crucial role in many biological processes. The interactions sites are often large and complex, therefore their mimicry by small molecules is a challenging task. An alternative approach includes application of peptides, which benefit from close structural resemblance to proteins and have enormous potential in the protein mimicry. Since discontinuous epitopes interactions with proteins play a major role in many diseases, inhibition or controlled modulation of PPIs participating in their development has become an important target in drug design. Despite efforts dedicated to the development of an HIV-1 vaccine, a successful candidate has not yet emerged. Among many evasive mechanisms of HIV, its error-prone replication and capability of mutation accumulation and emergence of mutants made the design of a successful vaccine an extremely difficult task. However, in order to maintain binding towards CD4 receptor, certain crucial residues within HIV must be conserved, which makes it an interesting approach towards development of the vaccine.

The main objective of this thesis was the synthesis of HIV gp120 protein mimics and evaluation of their binding to CD4 receptor.

In order to evaluate binding of the protein mimics, CD4 and gp120 proteins were needed. In chapter 2, efforts towards expression, purification and characterisation of CD4 and gp120 proteins are described. Both proteins were expressed successfully. However, the refolding of CD4D12 proved to be challenging and obtained protein was a mixture of properly folded and misfolded variants. gp120 was also expressed and purified, however low-yield did not allow to use this protein for evaluation assays and its further expression was discontinued. Future work should include further optimisation of refolding conditions of CD4D12 and scale-up of gp120 expression.

Synthesis of peptide-based mimics of gp120 conserved CD4-binding site was described in chapter 3. Three peptides corresponding to the gp120 discontinuous epitope were synthesized and cyclised in order to improve the mimicry of loop-like epitope structure and stability. The cyclisation was performed using a novel polar hinge, which improved their solubility. These peptides were mounted then on the TAC scaffold, which was used as a skeleton to display the peptide loops in

the correct 3D-conformation. Moreover, first approaches towards characterisation of the secondary structure of the synthesized linear and cyclic peptides as well as the final gp120 mimics by means of  $^1\text{H}$  NMR and CD-spectrometry are described. These techniques allowed to gain insight into the secondary structure features of the synthesized compounds and might aid future design of the constructs in order to make constructs biologically active.

Chapter 4 describes the development of a reliable and reproducible SPR method to evaluate binding of gp120 mimics to CD4 receptor. Careful optimisation of the experimental conditions yielded a highly reproducible method for the evaluation of gp120 discontinuous mimics. All tested constructs as well as loops 1 and 3 proved to be inactive or very weakly active, while loop 2 showed significant binding. The observed weak binding was probably due to an increased flexibility of the linker used for cyclisation of the peptides or peptide sequence specificity. It underlined the necessity to design a cyclisation linker, which would not only increase the solubility of the peptides but also increase their rigidity. Moreover, exploration scrambled peptide sequences and use of other molecular scaffolds which are more pre-organised could be possibly a remedy and give insight into to the low activity of the synthesized gp120 mimics.

## 6 Materials and methods

### 6.1 Molecular biology

DNA plasmids V1Jnstpagp120 and pET28CD4 were kindly donated by Prof. Raghavan Varadarajan.

#### 6.1.1 Plasmid purification

DNA plasmids were purified from *E. coli* cultures (DH5 $\alpha$  cells) using Wizard Plus SV Miniprep Kit (Promega, #A1330). Purifications were carried out in accordance to the supplied centrifugation protocol. DNA was eluted in Nuclease-free water.

#### 6.1.2 DNA quantification

DNA solutions were quantified by measuring the absorbance at 260 nm ( $A_{260}$ ) using the NanoDrop 2000C spectrophotometer (Thermo Scientific).

#### 6.1.3 DNA restriction digest

Restriction digests were set-up following supplied protocols. A typical restriction digest was set-up as follows: 1 $\times$  CutSmart buffer (5  $\mu$ L), DNA (1  $\mu$ g), enzyme Bln I (1  $\mu$ L), enzyme Sma I or SspI-HF (1  $\mu$ L) were mixed together and topped up to 50  $\mu$ L with sterile, deionised water. Enzyme Bln I (New England BioLabs, #R0585S) was used for both expressed proteins, enzymes Sma I (New England BioLabs, #R0141S) and SspI-HF (New England BioLabs, #R3132S) were used for CD4D12 and gp120 respectively. Digests were then left at 37 $^{\circ}$ C for 1 h. Restriction digests were then analysed by DNA electrophoresis.

#### 6.1.4 DNA electrophoresis

Certified<sup>TM</sup> Molecular Biology Agarose (Bio-Rad, #161-3101) at 1% (w/v) was dissolved in TBE buffer (90 mM Tris, 90 mM boric acid, 2.5 mM EDTA). The slurry was heated in a microwave until the agarose melted, then it was left until it was cool enough to handle. 10,000  $\times$  SYBR safe stain (Invitrogen<sup>TM</sup>, #S33102) concentrate was diluted 1:10,000 in agarose gel buffer. The molten agarose was poured into the cast and allowed to set at room temperature. Once the gel had set, the comb was removed and mounted in an electrophoresis tank filled with 1 $\times$

TBE buffer. The DNA samples were mixed with the gel loading dye Purple (6×, New England Biolabs, #B7024S) and loaded slowly into the wells alongside DNA molecular weight marker (1 Kb Plus DNA ladder, Invitrogen™, #10787018). Electrophoresis was carried out at 80 V for 55 min. The gels were visualised and imaged using myECL™ imager (Thermo Scientific).

### **6.1.5 Transformations**

Transformations were carried out in accordance with the protocol supplied with the MAX Efficiency® DH5α™ competent cells (Invitrogen, #18258-012). Briefly, a typical transformation was carried out as follows: 50 µL of competent cells were taken from -85°C storage and defrosted on ice. 1 ng of plasmid was added to the competent cells and mixed by gentle swirling. The cells were incubated for 30 min on ice, before heating for 45 s at 42°C in a water bath. The cells were placed back on ice for 2 min and 950 µL of S.O.C. medium (Invitrogen™, 15544-034) was added. The transformation mixture was incubated at 37°C with shaking (225 rpm) for 1 h. Then, 100 µL of transformations was spread onto kanamycin resistant agar plate. The plates were incubated overnight at 37°C.

### **6.1.6 Recipes**

#### **6.1.6.1 LB agar plates**

For 500 mL mixture tryptone (5 g), yeast extract (2.5 g), NaCl (5 g) and bacto-agar (7.5 g) were weighed out. Tryptone, yeast extract and NaCl were dissolved in 400 mL ddH<sub>2</sub>O and pH adjusted to 7.5. Bacto-agar was added, the mixture was made up to 500 mL with ddH<sub>2</sub>O and autoclaved.

## **6.2 Protein expression in Bacteria and Purification**

### **6.2.1 Protein Over-Expression**

Expression media (10 mL) was treated with selective antibiotics (at 1:1000 dilution from 50 mg·mL<sup>-1</sup> stock concentrations), warmed to 37°C and inoculated with a single colony from the streaked LB agar plate. The culture was incubated at 37°C overnight in a shaking incubator (200 rpm). After incubation, the bacterial growth was visually assessed by a cloudy haze in the media. The primary inoculum was

transferred into 500 mL of media containing selective antibiotics and incubated at 37°C, with 200 rpm shaking until reaching an OD<sub>600</sub> ≈ 0.8. At this point expression was induced with Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Fisher Scientific, #BP1755-10) at a final concentration of 1 mM. The cultures were left to express overnight at 37°C with shaking (200 rpm).

## **6.2.2 Recipes**

### **6.2.2.1 LB medium**

For 1 L mixture tryptone (10 g), yeast extract (5 g) and NaCl (10 g) were weighed out and dissolve in 800 mL ddH<sub>2</sub>O. pH was adjusted to 7.5, the mixture was made up to 1 L with ddH<sub>2</sub>O and autoclaved.

### **6.2.2.2 LB overnight express medium.**

For 250 mL mixture Overnight Express Instant LB Medium (11.25 g) was weighed out and dissolved in 250 mL ddH<sub>2</sub>O. 2.5 mL glycerol was added, and the mixture was gently swirled until the medium was dissolved. The medium was heated in a microwave oven on a high-power setting until appearance of bubbles. Then, the mixture was cooled to room temperature.

### **6.2.2.3 2×YT medium**

For 600 mL mixture tryptone (9.6 g), yeast extract (6 g) and NaCl (3 g) were weighed out and dissolved in 500 mL ddH<sub>2</sub>O. pH was adjusted to 7.0, the mixture was made up to 600 mL with ddH<sub>2</sub>O and autoclaved.

## **6.2.3 Cell Lysis - sonication**

Cells were harvested by centrifugation (9 000 × g, 30 min, 4°C), supernatant was decanted, and cells were resuspended in the lysis buffer (30 mL per 1 L of culture, 1×PBS + protease inhibitor (Thermo Scientific, #88666, 1 tablet per 10 mL of extract). Sonication was carried out on a Sanyo Soniprep 150 at 8 amplitude microns for 30 min of 15 s on, 15 s off, while keeping cells on ice. The lysate was centrifuged at 40 000 × g for 30 min at 4°C. Supernatant was decanted. A pale, white pellet of inclusion bodies was resuspended and washed with a solution of methionine (a few grams per litre of water)/1 mM EDTA. Then the inclusion bodies

were centrifuged ( $40\,000 \times g$ , 30 min,  $4^{\circ}\text{C}$ ) and the pellet was washed with ddH<sub>2</sub>O and centrifuged to remove EDTA ( $18\,000 \times g$ , 10 min,  $4^{\circ}\text{C}$ ,  $\times 3$ ).

#### 6.2.4 Ni-affinity Chromatography

The inclusion body pellet was resuspended in 50 mL of the solubilization buffer (6 M Gdn×HCl, 5 mM imidazole, 2 mM β-mercaptoethanol, PBS, pH 8.0) while mixing with a rotating wheel overnight. The lysate was filtered through a syringe filter 0.2 μm (sartorius stedim biotech, Minisart®, #16534-K) to remove any insoluble debris. The column (HiFliQ5 Ni-NTA column, Generon) was washed with 5 column volumes (CV) of ddH<sub>2</sub>O at  $1\text{ mL}\cdot\text{min}^{-1}$  to remove EtOH, then it was equilibrated with the solubilization buffer. The lysate was then loaded onto the resin, which was followed by the washing step with 10 CVs of the wash buffer (6 M Gdn×HCl, 20 mM imidazole, 2 mM β-mercaptoethanol, PBS, pH = 8.0). Finally, the bound protein was eluted with an increasing gradient from 20 mM imidazole to 300 mM imidazole over 6 CVs with the elution buffer (6 M Gdn×HCl, 300 mM imidazole, 2 mM β-mercaptoethanol, PBS, pH 8.0).

#### 6.2.5 Dialysis and refolding

The purified protein was immediately after purification dialyzed overnight at  $4^{\circ}\text{C}$  against 1 L of refolding buffer 1 (3 M Gdn×HCl, 10% sucrose, 1 mM EDTA, 1 mM L-Glutathione reduced (GSH, Sigma, #G4251), 0.1 mM L-glutathione oxidised (GSSG, Sigma, #G4376), pH 9.6) and then changed into 1 L of refolding buffer 2 (50 mM Na<sub>2</sub>CO<sub>3</sub>, 10% sucrose, 1 mM EDTA, 0.1 mM GSH, 0.01 mM GSSG, pH 9.6) for a second overnight dialysis at  $4^{\circ}\text{C}$ . Afterwards, the protein was exhaustively dialyzed against PBS, (pH 7.4) at  $4^{\circ}\text{C}$ , the buffer was exchanged for a fresh one every 4 h for the first two dialyses, and then overnight for the final dialysis.

#### 6.2.6 Protein sample concentration

The recovered fractions of the protein were filtered through 0.2 μm filter (sartorius stedim biotech, Minisart®, #16534-K) to remove any precipitated protein and concentrated using ultrafiltration tubes (VivaSpin 20 MWCO 10 kDa, GE Healthcare #28-0323-60). Columns were first pre-rinsed with ddH<sub>2</sub>O to remove any trace contaminants. Proteins samples were centrifuged at  $8000 \times g$  until the desired concentration was achieved. The concentrated pure protein sample was

aliquoted, a sample of each fraction was kept for Bradford assay and SDS-PAGE, then the rest was snap-frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ . Using this method, 2-3 mg of purified recombinant protein was obtained (per 1 L of bacterial cell culture) and 6.8 mg of protein that would require further purification.

## 6.2.7 Recipes

### 6.2.7.1 Preparation of dialysis tubing

For 2 L of 0.2 M  $\text{NaHCO}_3$ /5 mM EDTA solution  $\text{NaHCO}_3$  (33.6 g) and  $\text{EDTANa}_2$  (3.72 g) were weighed out.  $\text{NaHCO}_3$  and  $\text{EDTANa}_2$  were dissolved in 2 L of ddH<sub>2</sub>O. The required length of dialysis membrane (Visking Dialysis Tubing, Medicell, MWCO - 10-14000 Da, #DTV.12000.01) was cut. 1 L of 0.2 M  $\text{NaHCO}_3$ /5 mM EDTA solution was poured into a glass beaker, the membrane was inserted, and the solution was stirred and brought to boil. Afterwards, the solution was discarded, the membrane washed with ddH<sub>2</sub>O (2-3 $\times$ ), and the excess water poured off. The final 1 L of 0.2 M  $\text{NaHCO}_3$ /5 mM EDTA solution was added to cover the membrane and brought to boil. Then, the solution was discarded, the membrane was washed with ddH<sub>2</sub>O (2-3 $\times$ ), inserted in ddH<sub>2</sub>O and autoclaved. Prepared dialysis tubing was stored in ddH<sub>2</sub>O at  $4^{\circ}\text{C}$ .

## 6.3 Protein expression in mammalian cells and purification

### 6.3.1 Transfection and expression

Transfections were carried out in accordance to the protocol supplied with jetPRIME® reagent (Polyplus transfection, #114-01). Briefly, a typical transfection was carried out as follows:  $2 \times 10^6$  -  $5 \times 10^6$  HEK 293 cells were seeded in a 150 cm<sup>2</sup> flask 24 h prior to transfection. 10  $\mu\text{g}$  DNA was diluted in 1000  $\mu\text{L}$  pf jetPRIME™ buffer and mixed by vortexing. 20  $\mu\text{L}$  of jetPRIME™ reagent was added, the mixture was vortexed for 10 s, spinned down briefly and incubated for 10 min at room temperature. 1 mL pf transfection mix was added dropwise onto cells in regular cell growth medium and distributed evenly by gentle rocking plates back and forth and from side to side. The cultures were incubated at  $37^{\circ}\text{C}$  for 72 h.

As a transfection control, EGFP plasmid was also transfected into HEK293 cells using the same transfection conditions. The plasmid coding for EGFP was donated by Prof. Graeme Milligan (Institute of Molecular, Cell and Systems Biology, University of Glasgow). The fluorescence was measured using the excitation at 488 nm ( $\pm$  14 nm) and detection at 535 ( $\pm$  30 nm). The fluorescence was measured after 24 h, and at least a 2-fold increase in fluorescence was observed, which confirmed a successful transfection.

### 6.3.2 Purification

Purification was performed according to a slightly modified literature method.<sup>[139]</sup>

The tissue culture supernatant was collected and centrifuged (3000  $\times$  g, 20 min, 4°C). Empigen® BB detergent (Sigma, #30326) was added to a final concentration of 0.25% and the sample was directly used for purification.

The sample was loaded at 1 mL $\cdot$ min<sup>-1</sup> onto a 5 mL HiTrap™ Capto™ Lentil Lectin column (GE Healthcare, #17-5489-11) pre-equilibrated with binding buffer (150 mM NaCl, 20 mM Tris-HCl, 0.25% Empigen® BB, pH 7.5) which was followed by washing step with 10 CVs of high salt buffer (500 mM NaCl, 20 mM Tris-HCl, 0.25% Empigen® BB, pH 7.5). Then the column was washed with 10 CVs of binding buffer, and the protein was eluted with elution buffer (150 mM NaCl, 20 mM Tris-HCl, 0.25% Empigen® BB, 1 M MMP, pH 7.5) containing methyl- $\alpha$ -mannopyranoside (MMP, Sigma, #M6882).

### 6.3.3 Protein sample concentration

Protein containing fractions were pooled together and concentrated using Amicon® Ultra Centrifugal Filters (0.5 mL, 50K MWCO, Merck Millipore, #UFC505024). Columns were first pre-rinsed with ddH<sub>2</sub>O to remove any trace contaminants. Proteins samples were centrifuged at 14000  $\times$  g at 4°C until the desired concentration was achieved. The concentrated pure protein sample was aliquoted, a sample of each fraction was kept for Bradford assay and SDS-PAGE, then the rest was snap-frozen in liquid nitrogen and stored at -80°C.



## 6.3.4 Recipes

### 6.3.4.1 DMEM medium supplementation

50 mL of DMEM medium (Sigma, #D5671) was removed from the original 500 mL bottle. 50 mL of filter-sterilised Fetal Bovine Serum (Sigma, #F9665) was added directly to the bottle with DMEM. This was followed by addition of 5 mL of sterile Penicillin-Streptomycin mix (Sigma, #P0781) and 5 mL of GlutaMAX™ (Gibco®, #35050-038). Ready-to-use medium was stored at 2-8 °C

## 6.4 Protein Characterisation

### 6.4.1 Bradford assay

Coomassie Protein Assay Reagent (Thermo Scientific, #1856209) was removed from the 4 °C fridge and left to warm up to ambient temperature. A series of BSA (HyClone™ BSA, #SH40015.01) standards diluted with ddH<sub>2</sub>O to final concentrations of 0 (blank = no protein), 200, 400, 600, 800 and 1000 µg·min<sup>-1</sup> were prepared. A serial dilutions of the protein samples were prepared. Out of each sample 5 µL was loaded into a well of 96-well plate (each sample in triplicate). 200 µL of the dye reagent was added into each sample and the samples were incubated for 10 min with mixing using a Bibby Stuart GYRO-ROCKER® STR9. Absorbance was measured at 595 nm using microplate reader (CLARIOstar microplate reader with Corning® 96 well UV-transparent plates).

### 6.4.2 SDS-PAGE

Protein electrophoresis was carried out using Invitrogen's Mini Gel Tank system (Invitrogen, #A25977). The samples conditions were prepared by mixing 2.5 µg of the protein with 1.5 µL of 10× Bolt™ Sample Reducing Agent 10× (Invitrogen, #B0009), 4× loading buffer (Invitrogen, #B0007) and adding ddH<sub>2</sub>O to top up to 15 µL. The samples for non-reducing conditions were prepared in a similar way apart from excluding the reducing agent. Once prepared, the samples were heated at 70 °C for 10 min to aid unfolding. Then the samples alongside Spectra™ Multicolor Broad Range Protein Ladder (Thermo Fisher, #26634) were loaded onto Bolt™ 4-12% Bis Tris Plus gels (Invitrogen, #NW04125BOX) in 1× MES running buffer (20× Bolt™ MES SDS Running Buffer, Invitrogen, #B0002) and the gel was ran at

200V for 35 min. Gels were then washed with ddH<sub>2</sub>O for 5 min and fixed for 1 h at room temperature in a gel-fixing solution (50% (v/v) methanol in water with 10% (v/v) acetic acid). Then, the gels were washed with ddH<sub>2</sub>O (5 min, 5×) and stained with GelCode™ Blue Stain Reagent (Thermo Fisher, #24590) for 1 h at room temperature. Afterwards, the gels were destained in ddH<sub>2</sub>O overnight at room temperature.

### 6.4.3 Western Blot

Preceding western blotting, protein samples and pre-stained marker were separated by mass by SDS-PAGE. The protein bands from the unstained gels were transferred to PVDF Blotting Membrane (GE Healthcare, 0,2 µm, #10600021) using Mini Bolt™ module (Invitrogen, #B1000) following the supplied protocol. Western transfer was carried out in the transfer buffer (Bolt™ Transfer buffer, 20×, Invitrogen, #BT0006) at a constant voltage of 20 V for 1 h. Then the membrane was washed with 1× TBST (3× ,5 min) and blocked with a solution of 5% dried skimmed milk (Marvel) in 1× TBST buffer for 1 h at room temperature using Bibby Stuart GYRO-ROCKER® STR9. The membrane was then incubated with the primary antibody with gentle agitation overnight at 4 °C. The membrane was washed with 1× TBST (3× ,5 min), which was followed by an incubation with a secondary antibody for 1 h at room temperature. The membrane was washed with 1× TBST (3× ,5 min) and drained. Detection reagents (Pierce™ ECL Western Blotting Substrate, Thermo Scientific, #32106) were mixed at a 1:1 ratio, added to the blot and incubated for 1 min. The excess of the reagent was drained, the blot was covered with a clear plastic sheet protector and visualised using myECL™ imager (Thermo Scientific).

Antibodies used:

Rabbit Monoclonal Primary Anti-CD4 antibody [EPR7276], Abcam, #ab133622

Secondary antibody Goat anti-Rabbit IgG, Invitrogen, #31462

Primary Goat pAb to hiv1 gp120 abcam #ab85054 100ug 1mg/ml

Secondary antibody Donkey anti-Goat IgG (H+L), Invitrogen, #A16005

#### 6.4.4 Dot Blot

As a quick way to check in which fractions after purification the protein was present a dot blot procedure was carried out. On a Nitrocellulose Blotting membrane (GE HEALTHCARE, #10600002) a table of 2×2 cm was drawn. 5 µL sample of each fraction were pipetted in the centre of each square. The membrane was left until the samples were dry and then it was washed with 1× TBST (1× ,5 min) and blocked with a solution of 5% dried skimmed milk (Marvel) in 1× TBST buffer for 1 h at room temperature using Bibby Stuart GYRO-ROCKER® STR9. The membrane was then incubated with the primary antibody with gentle agitation for 1 h at room temperature. The membrane was washed with 1× TBST (3× , 5 min), which was followed by an incubation with a secondary antibody for 1 h at room temperature. The membrane was washed with 1× TBST (3× ,5 min) and drained. Detection reagents (Pierce™ ECL Western Blotting Substrate, Thermo Scientific, #32106) were mixed at a 1:1 ratio, added to the blot and incubated for 1 min. The excess of the reagent was drained, the blot was covered with a clear plastic sheet protector and visualised using myECL™ imager (Thermo Scientific).

#### 6.4.5 In-gel trypsin digest for MS analysis

25 µL of the purified protein was mixed with 5 µL of 5× loading buffer. The sample was heated at 90°C for 10 min. Then the samples were removed from the heat source and 1.5 µL of 1 M DTT, 6 µL of 500 mM iodoacetamide (in 100 mM ammonium bicarbonate solution) were added for reduction and alkylation and the sample was incubated in the dark for 5 min. The sample was loaded onto 10% Mini Protean® TGX™ gel (Bio-Rad, #456-1033) and separated with 1× SDS/glycine running buffer at 150 V for 60 min. The gel was stained for 10 min with Coomassie Blue stain and destained for 20 min in a destaining solution, which was followed by washing the gel with ddH<sub>2</sub>O (5 min, 3×). A piece of the gel where the protein of interest was resolved was separated and transferred into an Eppendorf and cut into several pieces. The gel sample was washed with acetonitrile-ddH<sub>2</sub>O (1:1) mixture. The solution was discarded, and the gel sample was washed with 100 mM ammonium bicarbonate (30 min). Then the gel pieces were washed with shaking in 50% acetonitrile/100 mM ammonium bicarbonate solution (1 h). The solvents were discarded and 50 µL of acetonitrile to shrink gel pieces. After 10 min the solvent was removed, and gel pieces were dried in a dry bath (30-38°C). 50 µL of

0.2  $\mu\text{g}\cdot\mu\text{L}^{-1}$  sequencing grade modified Porcine Trypsin (Promega, #V111) in 25 mM ammonium bicarbonate was added to the gel pieces to rehydrate them. The protein was left to digest overnight at 37°C. 50  $\mu\text{L}$  of 50% acetonitrile was added to the digest and incubated for 20 min. The sample was then briefly centrifuged to pellet gel pieces. The supernatant was transferred to clean tubes. 15  $\mu\text{L}$  of 80% acetonitrile/0.1% formic acid in ddH<sub>2</sub>O was added to the gel pieces and incubated for 20 min. The sample was then briefly centrifuged, and the supernatant was transferred to the same tube as used for the first extract. The wash step with formic acid solution was repeated once again. The solvents were then removed from the combined extract using Speed-Vac. 100  $\mu\text{L}$  of 0.1% formic acid was added to the sample, which was followed by centrifugation to remove any solid particles that could interfere with analysis. The lysate was carefully transferred to a clean vial and analysed by the Ultimate 300 (Dionex) HPLC system with the MS analysis of peptide sequences.

## 6.4.6 Recipes

### 6.4.6.1 10× TBST buffer

For 1 L of buffer NaCl (116 g) and Tris (2.42 g) were weighed out and dissolved in 800 mL ddH<sub>2</sub>O. pH was adjusted to 7.5, Tween® 20 (10 mL) was added and the buffer was made up to 1 L with ddH<sub>2</sub>O.

### 6.4.6.2 Destaining solution

Mix ddH<sub>2</sub>O, methanol and acetic acid in a ratio of 50/40/10 (v/v/v).

## 6.5 Binding assays – Surface Plasmon Resonance (SPR)

Experiments were performed with BIACore X-100 instrument using Sensor Chip CM5 (BIACore, Uppsala, Sweden, #BR-1000-12) at 25°C.

### 6.5.1 pH scouting and pre-concentration studies – determination of coupling conditions

Ligand solutions (10  $\mu\text{g}\cdot\text{mL}^{-1}$ , 75  $\mu\text{L}$ ), were prepared in 10 mM acetate buffers at different pH values: 5.5, 5.0, 4.5, 4.0. The ligand solution was injected using a contact time of 120 s in 1× HBS-P (prepared from the 10× concentrated stock

solution, pH 7.4, GE Healthcare, #BR100827) as a running buffer, starting at the highest pH to reduce the risk of aggregation or precipitation of the ligand. The chip surface was regenerated with 50 mM NaOH after each injection of the ligand.

### 6.5.2 Immobilization

Ligand solution was prepared at  $10\ \mu\text{g}\cdot\text{mL}^{-1}$  or  $20\ \mu\text{g}\cdot\text{mL}^{-1}$  in 10 mM acetate buffer pH 5.0 (175  $\mu\text{L}$ ) for CD4D12 and gp120, respectively. The surface in the flow cell 2 was activated by injection of the mixture of following solutions (provided in the Amine Coupling Kit, GE Healthcare, #BR100050): EDC (0.4 M in water), NHS (0.1 M in water) for 7 min. This was followed by injection of the ligand until the desired response was reached. Then 1 M ethanolamine pH 8.5 (Amine Coupling Kit) was injected for 7 min to deactivate remaining active groups on the surface and remove non-covalently bound ligand. The flow rate of  $10\ \mu\text{L}\cdot\text{min}^{-1}$  was used.

A naked sensor surface in flow cell 1, which was activated and deactivated in the same way as in flow cell 2, served as a negative control for each binding interaction.

### 6.5.3 Regeneration scouting

To find optimal regeneration conditions the regeneration scouting procedure was carried out. A previously unused sensor surface, with a representative for experiments amount of immobilized ligand, was used. An analyte was prepared at the highest concentration that was planned to be used in the binding experiments. Scouting was performed by testing a few (3-5) repeated cycles of analyte binding and regeneration with each of the conditions. Scouting was started with the mildest conditions and progressed towards harsher conditions. The entire procedure was as follows: the analyte was injected for 60 s, which was followed by injection of regeneration solution for 30 s. The regeneration solutions that were tested included 10 mM glycine at a range of pHs: 3.0, 2.5, 2.0, 1.75, 1.5.

### 6.5.4 Binding assays

Sensor chip was primed with degassed and filtered running buffer. 5 start-up cycles, consisting of injections of the running buffer followed by regeneration solution, were performed to obtain a stable baseline. The analyte was prepared

in serial dilutions (0-100 nM gp120, 0-100  $\mu$ M gp120 mimics, 0-250 nM CD4D12) and applied to the sensor chip at a flow rate of 30  $\mu$ L $\cdot$ min<sup>-1</sup> allowing 120 s for association and 600 s for dissociation. Between sample applications the chip was regenerated with two injections of the regeneration buffer for 30 s each. During the experiment, the response obtained using reference surface and the response obtained from the 2-3 blank injections (zero analyte concentrations) was subtracted (double referencing). Each sample concentration was performed in a duplicate and each experiment was repeated in total two times. The results were plotted as RU versus time and analysed with the BIAcore X100-evaluation software 2.0.1 (BIAcore, Uppsala, Sweden). Binding of all the molecules was fitted a 1:1 Langmuir binding model describing the interaction of two molecules in a 1:1 complex.

Used solutions:

for protein-protein interactions	HBS-P: 0.01 M HEPES pH 7.4, 0.15 M NaCl, 0.005% v/v surfactant P20, GE Healthcare, #BR100368
for protein-gp120 mimics or independent loops interactions	PBS-P+: 0.2 M phosphate buffer with 0.27 mM KCl, 1.37 M NaCl and 0.5% surfactant P20, pH 7.4, GE Healthcare, #28995084
Regeneration solution	10 mM Glycine pH 1.75

## 6.6 Synthesis

### 6.6.1 General experimental

All reagents were obtained from commercial sources and used without further purification unless specified otherwise. Fmoc-amino acids were obtained from Activotec (Cambridge, United Kingdom) and N,N,N',N'-Tetramethyl-O-(6-chloro-1H-benzotriazol-1-yl)uranium hexafluorophosphate (HCTU) was obtained from Matrix Innovation, Quebec, Canada). Tentagel S RAM resin (particle size 90  $\mu$ m, capacity 0.25 mmol $\cdot$ g<sup>-1</sup>) was obtained from IRIS Biotech (Markredwitz, Germany).

Methyl tert-butyl ether (MTBE), n-hexane (HPLC grade) and TFA were obtained from Aldrich (Milwaukee, USA). DMF (peptide grade) was obtained from VWR (Lutterworth, United Kingdom). Piperidine and DiPEA were obtained from AGC Bioproducts (Hessle, United Kingdom), and 1,2-ethanedithiol (EDT) was obtained from Merck (Darmstadt, Germany). HPLC grade DCM and acetonitrile were obtained from Fischer Scientific (Loughborough, United Kingdom). Solid phase peptide synthesis (SPPS) was performed on a PTI Tribute-UV peptide synthesizer. Lyophilisations were performed on a Christ Alpha 2-4 LDplus freeze dryer.

Reactions were carried out at ambient temperature unless stated otherwise. Air and/or moisture sensitive reactions were performed in the glass apparatus dried with a heat-gun under vacuum and flushed with nitrogen. Solvents were evaporated under reduced pressure at 40°C. Dry solvents of tetrahydrofuran, dichloromethane and diethyl ether were purified using a Pure-Solv<sup>TM</sup> 500 Solvent Purification System. All other solvents of analytical or HPLC grade were used without any further purification. Ultra-pure water was obtained with the use of Sartorius Arium<sup>®</sup> Comfort I water purification system.

Thin Layer Chromatography (TLC) was performed using Merck silica gel glass 60 F<sub>254</sub>. TLC plates were visualised by fluorescent quenching at  $\lambda = 254$  nm with incident UV light produced from a UV Minerallight<sup>®</sup> lamp. TLC plates were also visualised by staining using the most appropriated solution: potassium permanganate (5 g potassium carbonate and 10 g potassium permanganate in 500 mL water), p-anisaldehyde (10 mL of concentrated sulphuric acid, 10 mL glacial acetic acid, 20 mL of p-anisaldehyde to 360 mL of absolute ethanol). Flash column chromatography was performed using Merck silica gel 60 (40-63  $\mu$ m) from Silicycle (Canada).

<sup>1</sup>H-NMR and <sup>13</sup>C-NMR spectra were recorded on a Bruker<sup>®</sup> Avance<sup>™</sup> III 400 MHz and 500 MHz spectrometers at ambient temperature. Chemical shifts are quoted in parts per million (ppm) downfield of TMS ( $\delta_{\text{TMS}} = 0$  ppm). Data are reported as follows: chemical shift in ppm relative to CDCl<sub>3</sub> (7.26) or trimethylsilane (TMS, 0.00 ppm), multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, b = broad), coupling constant(s) *J* (Hz), integration and assignment.

Analytical HPLC was carried out with a Shimadzu Prominence system consisting of a communication module (CBM-20A), autosampler (SIL-20HT), pump modules (LC-20AT), UV/VIS detector (SPD-20A) and system controller (Labsolutions V5.54 SP), with a Dr. Maisch GMBH ReproSil-Pur Gold 200 C18 column (200 Å, 5 µm, 250 × 4.6 mm). UV absorbtion was recorded at 214 and 254 nm, by use of a standard protocol: 100% buffer A [acetonitrile/H<sub>2</sub>O 5:95 with TFA (0.1%)] for 2 min followed by a linear gradient of buffer B [acetonitrile /H<sub>2</sub>O 95:5 with TFA (0.1%)] to buffer A (0-100%) over 30 min at a flow rate of 1.0 mL min<sup>-1</sup>.

Purification of peptidic compounds was performed on an Agilent Technologies 1260 infinity preparative system coupled to a UV detector at 214 nm with a Dr. Maisch GMBH Reprosil Gold 200 C18 column (200 Å, 10 µm, 250 × 20 mm). Auto-collection of fractions was based on the UV measurements at 214 nm, using either 100% buffer A or 95% buffer A with 5% buffer B for 5 min followed by a linear gradient of buffer B into buffer A (specified for each compound) over 65 min at a flow rate of 12.5 mL·min<sup>-1</sup> using the same buffers as described for analytical HPLC.

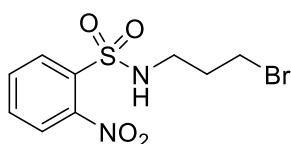
Liquid chromatography mass spectrometry (LC-MS) was carried out on a Thermo Scientific LCQ Fleet Quadrupole mass spectrometer and a Dionex Ultimate 3000 LC with use of a Dr. Maisch Reprosil Gold 120 C18 column (110 Å, 3 µm, 150 × 4.0 mm), and a 0-100% linear gradient of buffer B into buffer A with a flow of 1 mL·min<sup>-1</sup>.

High-resolution electrospray ionization (ESI) mass spectra were measured with a Bruker micrOTOF-Q II instrument in a positive mode.

### 6.6.2 Scaffold synthesis

Apart from compound **2**, the synthesis of TAC scaffold was performed according to the literature procedure.<sup>[88]</sup>

#### N-(3-bromopropyl)-2-nitrobenzene-sulfonamide (**7**)



Chemical Formula: C<sub>9</sub>H<sub>11</sub>BrN<sub>2</sub>O<sub>4</sub>S  
Molecular Weight: 323.16



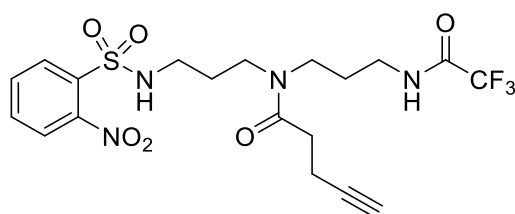
Bromopropylamine hydrobromide (43.6 g, 200 mmol) and 2-nitrobenzenesulfonyl chloride (53.2 g, 240 mmol) were dissolved in DCM (320 mL) and cooled in an ice bath. After dropwise addition of triethylamine (64.1 mL, 460 mmol) the reaction mixture was stirred overnight at room temperature affording a yellow suspension. The solvent was removed under vacuum and EtOAc (500 mL) was added and the mixture was washed with KHSO<sub>4</sub> solution (1 M, 250 mL, 2×), NaHCO<sub>3</sub> solution (1 M, 250 mL, 2×) and brine (250 mL, 2×). The organic phase was dried over MgSO<sub>4</sub>, filtered and the solvent was removed by evaporation and co-evaporation with CHCl<sub>3</sub> under vacuum, yielding the product as an off-white solid (61.68 g, 95%).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ = 2.11 (m, 2H), 3.28 (q, 2H), 3.47 (t, 2H, *J* = 6.2 Hz), 5.41 (t, 1H, *J* = 6.3 Hz), 7.77 (m, 2H), 7.89 (m, 1H), 8.16 (m, 1H).

<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>): δ = 29.90, 32.39, 41.97, 125.48, 131.11, 132.93, 133.41, 133.77, 148.04.

Spectroscopic data were in accordance with literature.<sup>[88]</sup>

***N*-(3-(2-nitrophenylsulfonamido)propyl)-*N*-(3-(2,2,2-trifluoroacetamido)propyl)pent-4-unamide (10)**



Chemical Formula: C<sub>19</sub>H<sub>23</sub>F<sub>3</sub>N<sub>4</sub>O<sub>6</sub>S  
Molecular Weight: 492.47

To a cooled solution (ice bath) of 1,3-diaminopropane (159 mL, 190.87 mmol) in DMA (650 mL) a solution of sulfonamide bromide **7** in DMA (650 mL) was added dropwise. The solution turned dark orange/red with a white precipitate. The resulting mixture was allowed to warm up to room temperature and was stirred overnight. An aqueous solution of 4 M NaOH (47.7 mL, 1 equiv.) was added after which the precipitate was fully dissolved. The mixture was concentrated in vacuo until a third of the volume. DMA (600 mL) was added and again the mixture was concentrated until a third of the volume remained. This co-evaporation was

repeated until the collected DMA was not basic (litmus paper). After a final evaporation of all remaining DMA, compound **8** was obtained as a yellow oil.

Acetonitrile (630 mL), H<sub>2</sub>O (3.4 mL, 190.9 mmol) and ethyl trifluoroacetate (90.85 mL, 763.5 mmol) were added to the crude intermediate **8**. The reaction was stirred overnight under reflux (90 °C) followed by concentration in vacuo to afford triamine **9** as a yellow oil.

Crude triamine compound **9** was dissolved in DCM (1200 mL). BOP (88.6 g, 200.4 mmol), 4-pentynoic acid (18.7 g, 190.9 mmol) and NMM (95.1 mL, 874.5 mmol) were added and the mixture was stirred overnight. The solvent was evaporated, and the residue was dissolved in EtOAc (600 mL) and divided into two portions. Each of the portions was washed with 1 M NaHCO<sub>3</sub> (400 mL, 2×), 1 M KHSO<sub>4</sub> (400 mL, 2×) and brine (400 mL). The organic layers were dried over MgSO<sub>4</sub>, filtered and the solvent was removed in vacuo. The crude product (99.43 g) was purified by column chromatography in a batch wise process with 10 to 15 g crude product until all purification was complete. For each batch, column chromatography was conducted with *ca.* 750 g silica and the gradient as follows: (EtOAc/hexane 4:6 until first a yellow band was eluted, then 1:1 until the product started to elute, then gradually increasing the amount of EtOAc up to 8:2 until the product had eluted) to obtain compound **10** as a yellow oil, which slowly solidified (65.12 g, 132 mmol, 69% over 3 steps). *R<sub>f</sub>* = 0.61 (EtOAc/hexane 8:2).

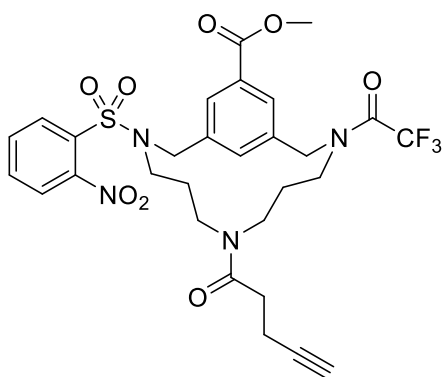
**<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):**  $\delta$  = 1.69-1.73, 1.85-1.90 (m, 4H, N-CH<sub>2</sub>-CH<sub>2</sub>), 1.96 (s, 1H, CCH), 2.56 (m, 4H, C(O)-CH<sub>2</sub>-CH<sub>2</sub>), 3.06-3.18, 3.25-3.30 (2m, 4H, 2 × NH-CH<sub>2</sub>-CH<sub>2</sub>), 3.41-3.44 (m, 4H, CH<sub>2</sub>-N-CH<sub>2</sub>), 5.58, 6.27, 6.90 (3m, 2H, NH), 7.71-7.90, 8.08-8.17 (2m, 4H, Ar-H).

**<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>):**  $\delta$  = 14.76, 14.90, 26.76, 28.07, 28.45, 29.01, 31.72, 31.85, 35.87, 37.40, 40.75, 41.05, 41.96, 42.21, 42.71, 45.01, 45.36, 69.01, 69.20, 82.98, 83.12, 114.58, 117.44, 125.16, 125.53, 130.69, 131.05, 132.72, 132.99, 133.17, 133.51, 133.98.

LCMS-ESI: *t<sub>R</sub>* = 16.59, exact mass calculated [M+H]<sup>+</sup>: 493.1290; found: 493.08.

Spectroscopic data in accordance with literature.<sup>[88]</sup>

## Triamine cyclization (12)



Chemical Formula: C<sub>29</sub>H<sub>31</sub>F<sub>3</sub>N<sub>4</sub>O<sub>8</sub>S  
Molecular Weight: 652.64

Dibromide **11** was provided by Dr Helmus van de Langemheen.

The reaction described below was performed ten times to react all of the starting material at the desired, high (12 mM) dilution.

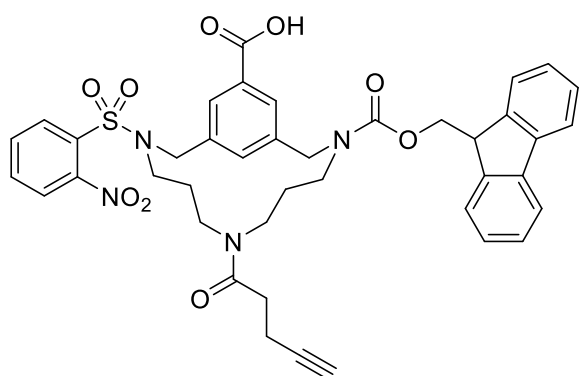
Compound **10** (6.5 g, 13.2 mmol, 1 equiv.), dibromide **11** (4.25 g, 13.2 mmol, 1 equiv.) and Cs<sub>2</sub>CO<sub>3</sub> (17.2 g, 52.8 mmol, 4 equiv.) were dissolved in DMF (1100 mL) and the resulting mixture was stirred overnight. The DMF was then evaporated in vacuo and EtOAc (650 mL) was added. The organic layer was washed with water (1 x 400 mL), 1 M KHSO<sub>4</sub> (1 x 450 mL) and with brine (1 x 450 mL), dried over MgSO<sub>4</sub> and filtered. Evaporation of the solvent afforded 85.08 g of the crude product (which was divided in several fractions for purification, each of them 15 - 18 g) as an orange oil which was purified using flash column chromatography (EtOAc/hexane 7:3). Compound **12** was obtained as a yellow foam (37.84 g, mmol, 44%). R<sub>f</sub> = 0.33 (EtOAc/hexane 7:3).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ = 1.26-1.42, 1.60-1.70 (2m, 4H, 2 × N-CH<sub>2</sub>-CH<sub>2</sub>), 1.92 (m, 1H, CCH), 2.38-2.45 (m, 4H, C(O)-CH<sub>2</sub>-CH<sub>2</sub>), 3.93, 3.96 (2s, 3H, OCH<sub>3</sub>), 4.41-4.52, 4.66-4.79 (2m, 4H, 2 × Ar-CH<sub>2</sub>-N), 7.67-8.10 (m, 7H, Ar-H).

LCMS-ESI: t<sub>R</sub> = 19.87, exact mass calculated [M+H]<sup>+</sup>: 653.1815; found: 653.08.

Spectroscopic data in accordance with literature.<sup>[88]</sup>

### Pentynoic acid amidated Fmoc protected TAC-scaffold (13)

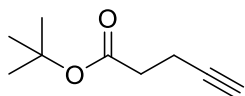


Chemical Formula: C<sub>41</sub>H<sub>40</sub>N<sub>4</sub>O<sub>9</sub>S  
Molecular Weight: 764.85

Compound 12 (13.0 g, 20 mmol) was dissolved in dioxane/MeOH/aq. NaOH (4 M) (15:4:1, 700 mL, 140 mmol) and the resulting mixture was stirred overnight at room temperature. 1 M HCl was added until the pH of the mixture was neutral, after which acetonitrile (500 mL) and H<sub>2</sub>O (500 mL) were added. The pH was adjusted to approximately 8 using DiPEA and a solution of Fmoc-OSu (7.4 g, 22 mmol, 1.1 equiv.) in Acetonitrile (75 mL) was added. This was followed by dropwise addition of DiPEA to maintain pH 8 and the reaction was considered to be complete when the pH remained above 7.5 for 10 min. For the purpose of ease of handling, the reaction mixture was divided into 3 equal fractions, each fraction was treated as follows: An aqueous solution of HCl (1 M, 75 mL), and H<sub>2</sub>O (500 mL) was added followed by extraction with EtOAc (2 × 500 mL). The combined organic layers were washed with brine (1 × 500 mL) and dried over MgSO<sub>4</sub>, filtered and evaporated in vacuo. Combination of all fractions afforded the crude product (17.9 g) as a yellow solid. Further purification by flash gel chromatography (MeOH/DCM 2:98 with 0.1% AcOH) gave the Fmoc-protected TAC scaffold 13 (14.6 g, 19 mmol, 95%). *R<sub>f</sub>* = 0.52 (6% MeOH/DCM with 0.1% AcOH).

LCMS-ESI: *t<sub>R</sub>* = 21.33, exact mass calculated [M+H]<sup>+</sup>: 765.2516; found: 765.00.

### Tert-butyl pent-4-ynoate (2)



Chemical Formula: C<sub>9</sub>H<sub>14</sub>O<sub>2</sub>  
Molecular Weight: 154.2063

4-pentynoic acid (5 g, 51 mmol) was dissolved in anhydrous DCM (75 mL) and an excess of tert-butyl-2,2,2-trichloroacetimidate (33.4 g, 153 mmol, 3 equiv.) was added. The reaction mixture was stirred overnight at room temperature, then it

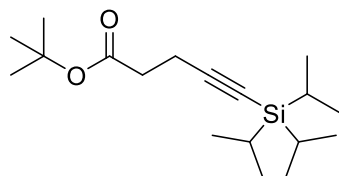
was filtered through celite to remove trichloroacetamide (white crystals). After filtration, the solvent was carefully removed by evaporation under reduced pressure with water bath kept at room temperature. Tert-butyl pent-4-ynoate was obtained as a yellow oil (7.1 g, 46 mmol, 90%).  $R_f = 0.43$  (Et<sub>2</sub>O/hexane 0.5/9.5)

<sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>):  $\delta$  = 1.45 (s, 9H, C(CH<sub>3</sub>)<sub>3</sub>), 1.96 (m, 1H, CH), 2.45 (m, 4H, CH<sub>2</sub>-CH<sub>2</sub>).

<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>):  $\delta$  = 14.45, 28.03, 34.46, 68.72, 80.72, 82.67, 170.96.

Spectroscopic data were in agreement with literature.<sup>[151]</sup>

#### Tert-butyl 5-(triisopropylsilyl)-pent-4-ynoate (4)



Chemical Formula: C<sub>18</sub>H<sub>34</sub>O<sub>2</sub>Si  
Molecular Weight: 310.5530

Compound 4 was prepared according to a modified literature procedure.<sup>[88]</sup>

Compound 2 (0.8 g, 5 mmol) was dissolved in anhydrous THF (15 mL) and cooled to -78°C (dry ice/acetone bath). *n*-BuLi (2.0 mL, 2.5 M in hexanes, 5 mmol) was added dropwise and the reaction was stirred for 30 min. the solution was warmed to 0°C by replacing the dry ice/ acetone bath with an ice bath and TIPS-Cl (1.3 mL, 6 mmol, 1.2 equiv.) was added dropwise. The reaction was stirred for 3 h at room temperature after which it was quenched with 30 mL saturated aqueous NH<sub>4</sub>Cl. Et<sub>2</sub>O was carefully removed by evaporation under reduced pressure with water bath kept at room temperature. The aqueous slurry was diluted with H<sub>2</sub>O (40 mL) and the mixture was extracted with Et<sub>2</sub>O (3 × 50 mL) and the combined organic layers were dried over MgSO<sub>4</sub>. After filtration the solvent was evaporated under reduced pressure with room temperature water bath. The crude product was purified by flash column chromatography (1% Et<sub>2</sub>O/pentane) to give the desired product 4 as a yellow oil (0.89 g, 2.3 mmol, 46%)  $R_f = 0.51$  (2% Et<sub>2</sub>O/Petroleum ether with 0.1% of AcOH).

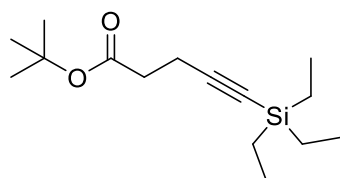
$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  = 1.04 (m, 21H,  $\text{Si}(\text{CH}(\text{CH}_3)_2)_3$ ), 1.44 (s, 9H,  $\text{C}(\text{CH}_3)_3$ ), 2.42-2.52 (m, 4H,  $\text{CH}_2\text{-CH}_2$ ).

$^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta$  = 11.22 ( $\text{SiCH}(\text{CH}_3)_2$ ), 15.92 ( $(\text{C}(\text{O})\text{CH}_2\text{CH}_2)$ ), 18.58 ( $\text{Si}(\text{CH}(\text{CH}_3)_2)_3$ ), 28.06 ( $\text{C}(\text{CH}_3)_3$ ), 35.14 ( $(\text{C}(\text{O})\text{CH}_2)$ ), 80.54, 80.77 ( $\text{CCH}_3$ ,  $\text{CCSi}$ ), 107.12 ( $\text{CCSi}$ ), 171.07 ( $\text{C=O}$ ).

HRMS-ESI: exact mass calculated  $[\text{M}+\text{Na}]^+$ : 333.2226; found: 333.2205.

Spectroscopic data was in agreement with literature <sup>[88]</sup>

### Tert-butyl 5-(triethylsilyl)-pent-4-ynoate (3)



Chemical Formula:  $\text{C}_{15}\text{H}_{28}\text{O}_2\text{Si}$   
Molecular Weight: 268.4720

Compound 3 was prepared according to a modified literature procedure.<sup>[88]</sup>

Compound 2 (2.7 g, 17 mmol) was dissolved in anhydrous THF (40 mL) and cooled to  $-78^\circ\text{C}$  (dry ice/acetone bath). *n*-BuLi (6.8 mL, 2.5 M in hexanes, 17 mmol) was added dropwise and the reaction mixture was stirred for 10 min. Dry ice/acetone bath was replaced by an ice bath and TES-Cl (3.4 mL, 20.4 mmol, 1.2 equiv.) was added dropwise. The reaction mixture was stirred for 3 h at room temperature, after which it was quenched with 50 mL of saturated aqueous  $\text{NH}_4\text{Cl}$ .  $\text{Et}_2\text{O}$  was carefully removed by evaporation under reduced pressure with a water bath at room temperature. The crude product was purified by flash column chromatography (1%  $\text{Et}_2\text{O}$ /pentane) to give the desired product as a yellow oil (1.96 g, 7.3 mmol, 43%)  $R_f$  = 0.53 (4%  $\text{Et}_2\text{O}$ /Petroleum ether with 0.1% AcOH).

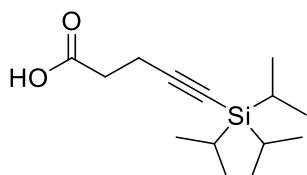
$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  = 0.55 (q,  $J$  = 7.8 Hz, 6H,  $\text{Si}(\text{CH}_2\text{-CH}_3)_3$ ), 0.96 (t,  $J$  = 7.9 Hz, 9H,  $\text{Si}(\text{CH}_2\text{-CH}_3)_3$ ), 1.44 (s, 9H,  $\text{CCH}_3$ ), 2.41-2.53 (m, 4H,  $\text{CH}_2\text{-CH}_2$ ).

$^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta$  = 4.45( $\text{Si}(\text{CH}_2\text{-CH}_3)_3$ ), 7.40 ( $\text{Si}(\text{CH}_2\text{-CH}_3)_3$ ), 15.96 ( $\text{C}(\text{O})\text{CH}_2\text{-CH}_2$ ), 28.05 ( $\text{CCH}_3$ ), 35.00 ( $\text{C}(\text{O})\text{CH}_2$ ), 80.59, 82.16 ( $\text{CCH}_3$ ,  $\text{CCSi}$ ), 106.49 ( $\text{CCSi}$ ), 171.09 ( $\text{C=O}$ ).

HRMS-ESI: exact mass calculated  $[M+Na]^+$ : 291.1756; found: 291.1754.

Spectroscopic data were in agreement with literature.<sup>[88]</sup>

### 5-(triisopropylsilyl)-4-pentynoic acid (**6**)



Chemical Formula:  $C_{14}H_{26}O_2Si$   
Molecular Weight: 254.4450

Compound **6** was prepared according to a modified literature procedure.<sup>[88]</sup>

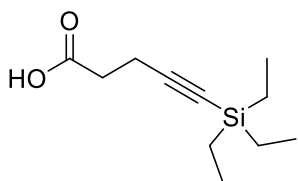
Compound **4** (1.3 g, 4.2 mmol) was dissolved in 10% TFA/DCM (50 mL) and the resulting mixture was stirred for 2 h. Then the reaction was quenched with 1 M ammonium acetate (100 mL) and extracted with DCM (2 × 75 mL). The combined organic layers were dried over  $MgSO_4$ , filtered and evaporated under reduced pressure. The crude product was purified using flash column chromatography (5-7%  $Et_2O$ /pentane with 0.1% AcOH). Eluted fractions were merged and solvents were evaporated under reduced pressure. Excess acetic acid was then removed by an additional washing step rather than by evaporation due to compound volatility as follows: The concentrated fractions were dissolved in  $Et_2O$  (15 mL) and washed with  $H_2O$  to remove AcOH (3 × 10 mL). Then, the combined organic layers were dried over  $MgSO_4$ , filtered and concentrated. The product **6** was obtained as a yellow oil (0.61 g, 2.4 mmol, 57%).  $R_f$  = 0.46 (20% EtOAc/hexanes with 0.1% AcOH).

$^1H$  NMR (400 MHz,  $CDCl_3$ ):  $\delta$  = 1.05 (m, 21H,  $Si(CH_2CH_2CH_3)_3$ ), 2.54-2.65 (m, 4H,  $\underline{CH_2-CH_2}$ ).

$^{13}C$  NMR (100 MHz,  $CDCl_3$ ):  $\delta$  = 11.18 ( $Si(\underline{CH_2-CH_3})_3$ ), 15.65 ( $C(O)CH_2-\underline{CH_2}$ ), 18.53 ( $Si(CH_2-\underline{CH_3})_3$ ), 33.80 ( $C(O)\underline{CH_2}$ ), 81.62 ( $\underline{CCSi}$ ), 106.21 ( $\underline{CCSi}$ ), 178.05 ( $C=O$ ).

HRMS-ESI: exact mass calculated  $[M+Na]^+$ : 277.1600; found: 277.1589.

Spectroscopic data were in agreement with literature.<sup>[88]</sup>

**5-(triethylsilyl)-4-pentynoic acid (5)**

Chemical Formula: C<sub>11</sub>H<sub>20</sub>O<sub>2</sub>Si  
Molecular Weight: 212.3640

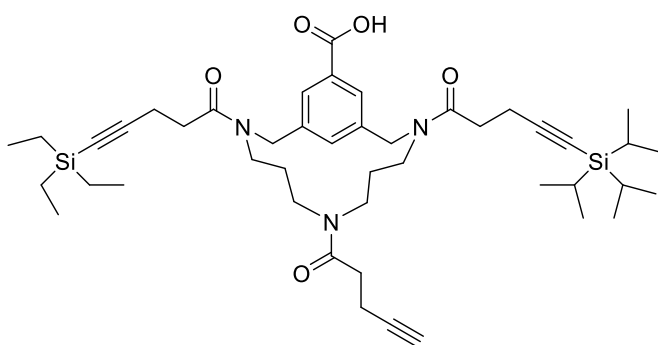
Compound **3** (0.81 g, 7.3 mmol) was dissolved in 15% TFA/DCM (50 mL) and the resulting mixture was stirred for 2 h. The reaction was then quenched with 1 M ammonium acetate (100 mL) and extracted with DCM (2 × 75 mL). The combined organic layers were dried over MgSO<sub>4</sub>, filtered and evaporated under reduced pressure. The crude product was purified by flash column chromatography (5-7% Et<sub>2</sub>O/pentane with 0.1% AcOH). Eluted fractions were merged and solvents were evaporated under reduced pressure. Afterwards, the crude product was dissolved in Et<sub>2</sub>O (15 mL) and washed with H<sub>2</sub>O to remove AcOH (3 × 10 mL). Then, the combined organic layers were dried over MgSO<sub>4</sub>, filtered and concentrated. The product **5** was obtained as a yellow oil (0.42 g, 1.97 mmol, 66%). *R<sub>f</sub>* = 0.43 (20% EtOAc/hexanes with 0.1% AcOH).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ = 0.56 (q, *J* = 7.9 Hz, 6H, Si(CH<sub>2</sub>-CH<sub>3</sub>)<sub>3</sub>), 0.96 (t, *J* = 7.9 Hz, 9H, Si(CH<sub>2</sub>-CH<sub>3</sub>)<sub>3</sub>), 2.53-2.64 (m, 4H, CH<sub>2</sub>-CH<sub>2</sub>).

<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>): δ = 4.40 (Si(CH<sub>2</sub>-CH<sub>3</sub>)<sub>3</sub>), 7.34 (Si(CH<sub>2</sub>-CH<sub>3</sub>)<sub>3</sub>), 15.61 (C(O)CH<sub>2</sub>-CH<sub>2</sub>), 33.58 (C(O)CH<sub>2</sub>), 82.69 (C≡Si), 105.86 (C≡Si), 176.86 (C=O).

HRMS-ESI: exact mass calculated [M+Na]<sup>+</sup>: 235.1130; found: 235.1120

Spectroscopic data were in agreement with literature.<sup>[88]</sup>

**Orthogonally protected trialkyne TAC-scaffold (15)**

Chemical Formula: C<sub>45</sub>H<sub>69</sub>N<sub>3</sub>O<sub>5</sub>Si<sub>2</sub>  
Molecular Weight: 788.23



TAC-scaffold **13** (0.45 g, 0.59 mmol) was loaded onto 2-chlorotriyl resin by dissolving in DCM (10 mL). DiPEA (102  $\mu$ L, 0.59 mmol) was added, followed by addition of 2-chlorotriyl chloride resin (1.0 g, 0.8 mmol). After 5 min, another portion of DiPEA (154  $\mu$ L, 0.88 mmol, 1.5 equiv.) was added and the mixture was stirred overnight at room temperature. DiPEA (1 mL) and MeOH (4 mL) were then added and the mixture was stirred for 30 min. The resin was transferred to a solid phase reaction vessel and washed with DCM (3  $\times$  20 mL), MeOH (3  $\times$  20 mL) and Et<sub>2</sub>O (3  $\times$  20 mL).

*Resin load determination.* After drying for 1 h, 8.9 mg of the loaded resin was transferred to a 25 mL volumetric flask and 20% piperidine/DMF (2 mL) was added. The flask was shaken for 30 min, after which MeOH was added until a volume of 25 mL. UV absorption was measured at 300 nm ( $\epsilon$  = 7800 M<sup>-1</sup>cm<sup>-1</sup>) resulting in a loading of 0.25 mmol·g<sup>-1</sup> resin (loading yield of 43%).

*Fmoc deprotection.* The rest of the resin was washed with DCM (3  $\times$  20 mL), DMF (3  $\times$  20 mL) and 20% piperidine/DMF (1  $\times$  20 mL) was added and the mixture was shaken for 1 min, then the solvents were discarded. Another 20 mL of 20% piperidine/DMF was added and the mixture was shaken for 30 min, after which the solvents were discarded. The resin was washed with DMF (3  $\times$  20 mL), DCM (3  $\times$  20 mL). A positive Bromophenol Blue-test indicated Fmoc removal.

*Coupling of TIPS-protected pentynoic acid 6.* BOP (0.22 g, 0.5 mmol, 2 equiv.), DMF (10 mL), TIPS protected pentynoic acid **6** (0.13 g, 0.5 mmol, 2 equiv.) DiPEA (0.17 mL, 1.0 mmol, 4 equiv.) were added. The mixture was shaken for 2 h. Then, the resin was washed with DMF (3  $\times$  20 mL), DCM (3  $\times$  20 mL). A negative Bromophenol Blue-test indicated coupling of the pentynoic acid derivative.

*oNBS deprotection.* The resin was washed with DMF (3  $\times$  20 mL). DMF (15 mL), B-mercaptoethanol (0.18 mL, 2.5 mmol, 10 equiv.) and DBU (0.19 mL, 1.3 mmol, 5 equiv.) were added and the mixture was shaken for 15 min. The deprotection step was repeated once and the resin was washed with DMF (3  $\times$  20 mL), DCM (3  $\times$  20 mL). A positive Bromophenol Blue-test indicated oNBS protecting group removal.

*Coupling of TES-protected pentynoic acid 5.* BOP (0.22 g, 0.5 mmol, 2 equiv.), DMF (10 mL), TES-protected pentynoic acid **5** (0.11 g, 0.5 mmol, 2 equiv.) and DiPEA (0.18 mL, 1.0 mmol, 4 equiv.) were added and the mixture was shaken for 2 h. The resin was washed with DMF (3 × 20 mL), DCM (3 × 20 mL) and a negative Bromophenol Blue-test indicated successful coupling.

*Resin cleavage.* The resin was transferred to a round bottom flask and a solution of 30% HFIP/DCM (20 mL) was added. The mixture was stirred for 45 min. After filtration and washing of the residue with DCM, EtOAc (30 mL) was added to the filtrate. The solvents were removed by evaporation under reduced pressure. Flash column chromatography (5% MeOH/DCM) afforded the scaffold as a yellow oil. The oil was dissolved in *t*-BuOH/H<sub>2</sub>O 1:1 and lyophilised to obtain a white solid **15** (0.17 g, 0.22 mmol, 88%). *R*<sub>f</sub> = 0.73 (10% MeOH/DCM). *t*<sub>R</sub> = 31.862,

LCMS-ESI: average mass calculated [M+H]<sup>+</sup>: 788.4854; found: 788.42, HRMS exact mass calculated [M+Na]<sup>+</sup>: 810.4673, found: 810.4628, MALDI-TOF found: 788.45.

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ = 0.57 (q, *J* = 7.9 Hz, 6H, Si(CH<sub>2</sub>CH<sub>3</sub>)<sub>3</sub>), 0.97 (t, *J* = 7.9 Hz, 9H, Si(CH<sub>2</sub>CH<sub>3</sub>)<sub>3</sub>), 1.05 (m, 21H, Si(CH(CH<sub>3</sub>)<sub>2</sub>)<sub>3</sub>), 1.25-1.56 (m, 4H, N-CH<sub>2</sub>-CH<sub>2</sub>-CH<sub>2</sub>-N), 1.94 (m, 1H, CCH), 2.45 (m, 4H, CH<sub>2</sub>CH<sub>2</sub>CCH), 2.60-2.79 (m, 8H, 2 × CH<sub>2</sub>CH<sub>2</sub>CCSi), 2.85-3.05, 3.40-3.50 (2m, 8H, N-CH<sub>2</sub>-CH<sub>2</sub>-CH<sub>2</sub>-N), 4.63-4.69 (m, 4H, 2 × N-CH<sub>2</sub>-Ar), 7.40-7.47, 7.89-7.95 (dd, *J* = 194.3, 27.6 Hz, 2H, Ar-H), 8.02-8.06 (d, 16 Hz, 1H, Ar-H).

<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>): δ = 4.46, 7.49, 11.26, 14.55, 14.60, 16.39, 16.41, 18.65, 28.00, 28.09, 28.36, 31.91, 31.93, 32.84, 33.11, 33.15, 43.55, 45.56, 45.61, 45.72, 46.04, 48.24, 28.30, 52.10, 52.18, 53.85, 53.92, 81.28, 81.34, 82.70, 82.75, 83.11, 83.22, 106.45, 106.48, 107.07, 107.15, 128.82, 130.03, 130.07, 130.09, 130.34, 131.37, 131.38, 131.41, 131.43, 138.19, 140.51, 168.85, 170.80, 172.12, 172.55.

Spectroscopic data were in agreement with literature.<sup>[88]</sup>

## 6.6.3 Peptide synthesis

### 6.6.3.1 Solid phase synthesis of linear peptides

*General method for automated peptide synthesis:* Linear peptides were synthesized on a PTI Tribute-UV peptide synthesizer. Tentagel S RAM resin (1.0 g, 0.25 mmol 1 equiv.) was allowed to swell ( $3 \times 10$  min). Removal of Fmoc-group was performed using 20% piperidine in DMF using Tribute-UV peptide synthesizer RV\_top\_UV\_Xtend protocol which was followed by a DMF washing step ( $5 \times 30$  s). Fmoc-protected amino acids (0.25 mmol, 4 equiv.) were coupled using HCTU (1 mmol, 4 equiv.) and DiPEA (2 mmol, 8 equiv.) in DMF as a solvent. The coupling time was 20 min. After each coupling the resin was washed with DMF ( $6 \times 30$  s). Capping was performed using a mixture of acetic anhydride (23.6 mL), DiPEA (10.9 mL) in DMF (500 mL). After the last amino acid coupling, the Fmoc group was cleaved under normal deprotection conditions (described above) and the resulting free N-terminus was acetylated by treatment of the resin bound peptide with acetic anhydride (250  $\mu$ L) and DiPEA (2 mmol, 8 equiv.) in DMF using coupling times described above. After the last coupling step, the resin was washed with DMF ( $5 \times 30$  s), DCM ( $5 \times 30$  s), dried over a nitrogen flow for 10 min, followed by the cleavage of the resin bound peptide.

Note: in the synthesis of linear loop 2 after the last Fmoc-removal the N-terminus was left as a free amine.

#### **General procedure for the cleavage and deprotection of the peptide from the solid support**

Deprotection of side chain protected amino acids and cleavage from the resin was achieved by treatment of the resin with a mixture of TFA:H<sub>2</sub>O:EDT:TIS (10 mL, 90/5/2.5/2.5, v/v/v/v) for 3 h. Afterwards, the peptide was precipitated with MTBE/hexane (90 mL, 1:1 v/v). After centrifugation (3500 rpm, 5 min) the supernatant was discarded and the pellet was suspended in MTBE:hexane (90 mL, 1:1, v/v) and centrifuged again. This procedure was repeated in total 3 times. Finally, the pellet was dissolved in *t*-BuOH/H<sub>2</sub>O (1:1, v/v) and lyophilized to yield the crude linear peptide.

A fraction of each of the crude linear peptides was purified for MALDI-TOF,  $^1\text{H}$ -NMR and CD-spectrometry purposes. Otherwise, all of the crude peptides were used for cyclization.

#### Linear peptides synthesized:

**Linear peptide 1**  $\times$  2TFA, Ac-CLTRDGGKC-NH<sub>2</sub>,  $t_R$  = 13.461, LCMS-ESI: average mass calculated  $[\text{M}]^+$ : 993.17; found: 993.58. MALDI-TOF: 993.32.

**Linear peptide 2**  $\times$  2TFA, H-CINMWQEVGKAC-NH<sub>2</sub>,  $t_R$  = 17.394, LCMS-ESI: average mass calculated  $[\text{M}]^+$ : 1380.67, found: 1380.92. MALDI-TOF: 1380.04.

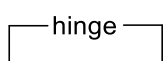
**Linear peptide 3**, Ac-CSGGDPEIVTC-NH<sub>2</sub>,  $t_R$  = 14.358, LCMS-ESI: average mass calculated  $[\text{M}]^+$ : 1121.25; found: 1121.08. MALDI TOF: as Na<sup>+</sup> adduct: 1143.23, 1165.07.

#### 6.6.3.2 Cyclic peptide synthesis

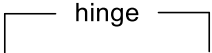
Cyclization hinge N<sub>3</sub>-TADB was provided by Dr Helmus van de Langemheen.

All cyclisations were performed at the concentration of 1 mM of the linear peptide. The crude linear peptide and 1,3,5-triazinane hinge N<sub>3</sub>-TADB were placed in a round bottomed flask and dissolved in acetonitrile. After 5 min, an aqueous solution of ammonium bicarbonate buffer (20 mM, pH = 7.8) was added to form a 1:3 v/v mixture of acetonitrile/buffer. The progress of the cyclization was checked by LCMS after 30 min, which confirmed reaction completion. Next, the solvents were removed in vacuo until approximately 1/3 of the volume, and the remaining slurry was lyophilised. The crude cyclic peptides were purified using preparative HPLC. The product containing fractions were pooled together and lyophilised to yield the desired cyclic peptides as a white fluffy powder.

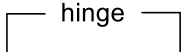
#### Cyclic peptides synthesized:

**Loop 1**  $\times$  2·TFA (16), , purified using gradient 0-40% of buffer B into buffer A over 40 min at a flow rate of 12.5 mL·min<sup>-1</sup>. The title compound was obtained as a white fluffy powder after lyophilisation (49.3 mg, 33.5  $\mu\text{mol}$ ,

13%). (21 steps, average 91% per step). aHPLC  $t_R$  = 14.593, LCMS-ESI:  $t_R$  = 8.67, average mass calculated  $[M]^+$ : 1243.38; found: 1243.25. MALDI-TOF found: 1243.23.


  
hinge

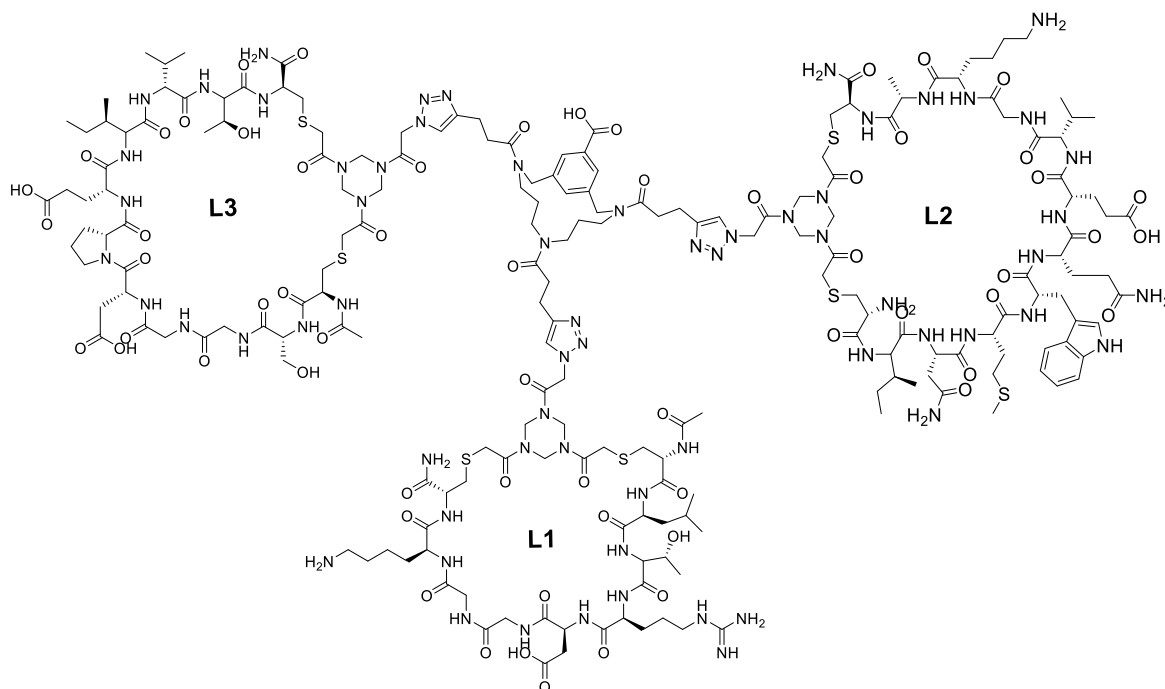
**Loop 2 × 2·TFA (17)**,  $\text{H-CINMWQEVGKAC-NH}_2$ , purified using gradient 5-40% of buffer B into buffer A over 35 min at a flow rate of  $12.5 \text{ mL} \cdot \text{min}^{-1}$ . The title compound was obtained as a white fluffy powder after lyophilisation: 47.6 mg (25.6  $\mu\text{mol}$ ), 10%, (26 steps, average 92% per step). aHPLC  $t_R$  = 17.023, LCMS-ESI:  $t_R$  = 11.88, average mass calculated  $[M+2H]^{2+}$ : 816.44; found: 816.08. Average mass calculated  $[M]^+$ : 1630.87, MALDI-TOF found: 1630.56.


  
hinge

**Loop 3 × 2·TFA (18)**,  $\text{Ac-CSGGDEPIVTC-NH}_2$ , purified using gradient 14-34% of buffer B into buffer A over 40 min at a flow rate of  $10 \text{ mL} \cdot \text{min}^{-1}$ . The title compound was obtained as a white fluffy powder after lyophilisation: 99.8 mg (72.8  $\mu\text{mol}$ ), 29%, (25 steps, average 95% per step). aHPLC  $t_R$  = 15.227, LCMS-ESI:  $t_R$  = 10.24, average mass calculated  $[M]^+$ : 1371.46.; found: 1371.50. Average mass calculated  $[M]^+$ : 1371.46, MALDI-TOF found: 1371.40.

### 6.6.3.3 Preparation of gp120 mimics

#### Synthesis gp120 mimic 22:



*First CuAAC (copper-catalysed azide-alkyne cycloaddition) and TES-removal:* TAC-scaffold **15** (5.12 mg, 6.5  $\mu\text{mol}$ , 1 equiv.), peptide loop **1** **16** (11.48 mg, 7.8  $\mu\text{mol}$ , 1.2 equiv.) and TBTA (1.72 mg, 3.25  $\mu\text{mol}$ , 0.5 equiv.) were placed in a flask and degassed DMF (300  $\mu\text{L}$ ) was added. The reaction was stirred under nitrogen atmosphere until all components were dissolved. Then, aqueous 1 M  $\text{CuSO}_4$  (3.5  $\mu\text{L}$ , 3.25  $\mu\text{mol}$ , 0.5 equiv.) and aqueous 0.25 M sodium ascorbate (38.6  $\mu\text{L}$ , 9.75 mmol, 1.5 equiv.) were added. The reaction mixture was stirred under nitrogen and monitored by analytical HPLC and LCMS. When reaction was complete (typically after 2 h),  $\text{AgNO}_3$  (15.6 mg, 91.83  $\mu\text{mol}$ , 14 equiv.) was added and the mixture was stirred under nitrogen atmosphere, and monitored by analytical HPLC and LCMS, which showed complete conversion after 1 h. Then, NaCl (5.5 mg, 91.83  $\mu\text{mol}$ , 14 equiv.) was added in order to precipitate out  $\text{Ag}^+$  as  $\text{AgCl}$ . Then, the solvents were removed with a nitrogen stream and the red residue was diluted with buffer A (1500  $\mu\text{L}$ ). The solution was centrifuged (4500 rpm, 5 min) and the supernatant was purified using a gradient of 30-70% of buffer B into buffer A over 40 min at a flow rate of 12.5  $\text{mL} \cdot \text{min}^{-1}$ . TES-protected product fractions were pooled and lyophilized to obtain a white fluffy solid of **19** (9.3 mg, 4.34  $\mu\text{mol}$ , overall yield: 67%, yield per step: 82%). aHPLC  $t_R$  = 24.246, LCMS-ESI:  $t_R$  = 18.78, average mass calculated  $[\text{M}+2\text{H}]^{2+}$ : 959.17; found: 959.42.

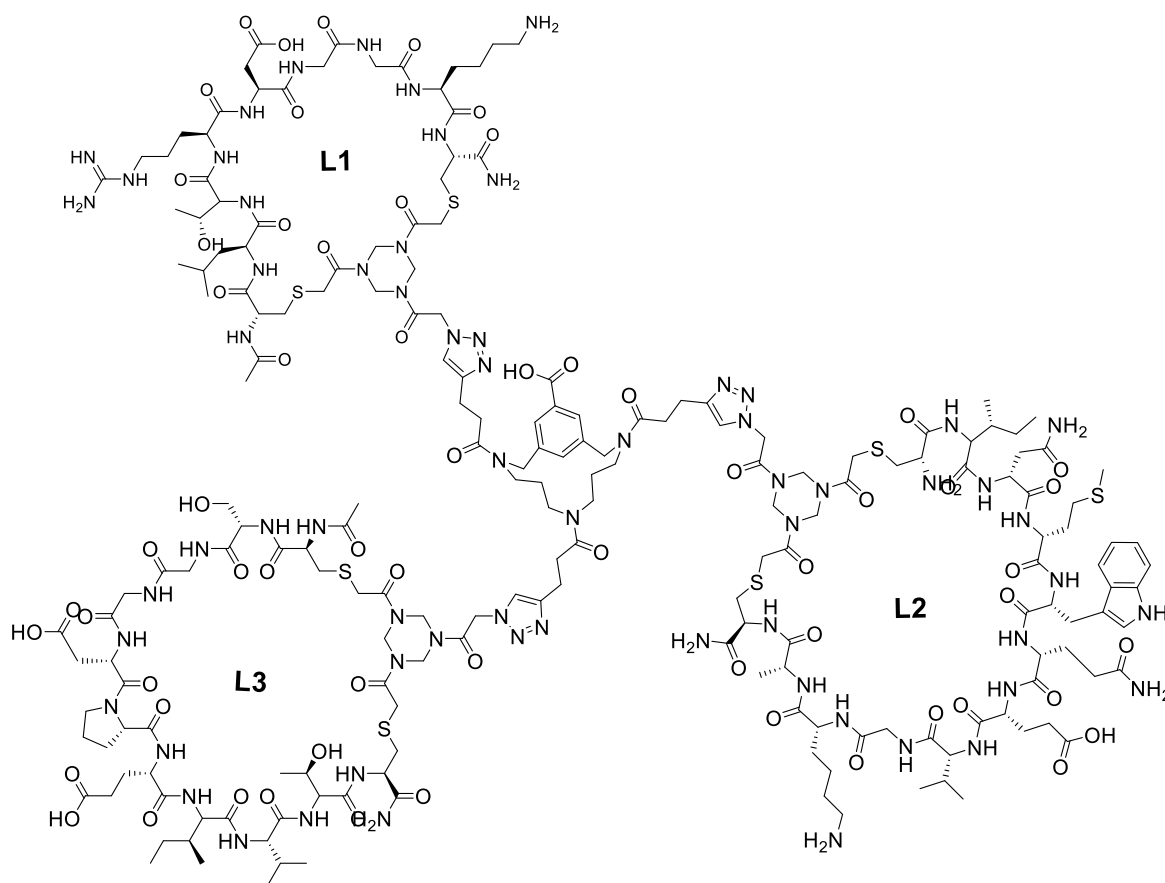
*Second CuAAC:* TES-protected TAC scaffold (**9**, 9.3 mg, 4.34  $\mu\text{mol}$ , 1 equiv.) peptide loop 3 **18** (7.14 mg, 5.21  $\mu\text{mol}$ , 1.2 equiv.), TBTA (0.7 mg, 1.3  $\mu\text{mol}$ , 0.3 equiv.), aqueous 1 M  $\text{CuSO}_4$  (6.1  $\mu\text{L}$ , 1.3  $\mu\text{mol}$ , 0.3 equiv.) and aqueous 0.25 M sodium ascorbate (20.6  $\mu\text{L}$ , 5.21  $\mu\text{mol}$ , 1.2 equiv.) were added and the resulting mixture was stirred under nitrogen atmosphere until all the reactants were dissolved. Then degassed deionized  $\text{H}_2\text{O}$  (155  $\mu\text{L}$ ) was added. The progress of the reaction was monitored by analytical HPLC and LCMS. After 4 h the reaction was complete, and solvents were removed by a nitrogen stream. The remaining residue was diluted with buffer A (900  $\mu\text{L}$ ) and the resulting solution was centrifuged (4500 rpm, 5 min) and purified using a gradient of 35-65% of buffer B into buffer A over 60 min at a flow rate of 12.5  $\text{mL}\cdot\text{min}^{-1}$ . The product containing fractions were pooled and lyophilized to obtain the scaffold derivative **20** with two peptides attached as a white fluffy solid (11.5 mg in 89% purity, 2.90  $\mu\text{mol}$ , yield 68%). aHPLC  $t_R = 21.787$ , LCMS-ESI:  $t_R = 16.78$ , average mass calculated  $[\text{M}+3\text{H}]^{3+}$ : 1097.27; found: 1097.08.

*TIPS removal:* TIPS-protected scaffold **20**  $\times$  2-TFA (10.2 mg, 2.90  $\mu\text{mol}$ ) was dissolved in degassed DMF (300  $\mu\text{L}$ ) and 1M TBAF in THF (42  $\mu\text{L}$ , 145  $\mu\text{mol}$ , 50 equiv.) was added. The reaction mixture was stirred under nitrogen atmosphere and monitored by analytical HPLC and LCMS. After 21 h the reaction was incomplete, hence another portion of 1 M TBAF in THF was added (21  $\mu\text{L}$ , 72.5  $\mu\text{mol}$ , 25 equiv.). After a further 5 h the reaction was still incomplete, hence another portion of 1 M TBAF in THF was added (10  $\mu\text{L}$ , 36.25  $\mu\text{mol}$ , 12.5 equiv.). Reaction completion was confirmed after a further 3.5 h and DMF was removed with nitrogen stream. The remaining solid was diluted with buffer A (350  $\mu\text{L}$ ) and centrifuged (4500 rpm, 5 min). The resulting solution was purified using a gradient of 0-55% of buffer B into buffer A over 55 min at 12.5  $\text{mL}\cdot\text{min}^{-1}$ . The TIPS-deprotected product was obtained as a white fluffy solid of **21** (5.8 mg, 1.73  $\mu\text{mol}$ , 60%) after lyophilization. aHPLC  $t_R = 16.294$ , LCMS-ESI:  $t_R = 11.16$ , average mass calculated  $[\text{M}+3\text{H}]^{3+}$ : 1044.82; found: 1044.83.

*Third CuCAAC:* TIPS-deprotected scaffold **21** (5.8 mg, 1.73  $\mu\text{mol}$ , 1 equiv.), peptide loop 2 **17** (3.21 mg, 1.73  $\mu\text{mol}$ , 1 equiv.), TBTA (0.46 mg, 0.86  $\mu\text{mol}$ , 0.5 equiv.) were placed in a vial and dissolved in degassed DMF (300  $\mu\text{L}$ ) and stirred under nitrogen atmosphere until dissolved. Then, aqueous 1 M  $\text{CuSO}_4$

(0.9  $\mu\text{L}$ , 0.9  $\mu\text{mol}$ , 1.1 equiv.) and aqueous 0.25 M sodium ascorbate (10.2  $\mu\text{L}$ , 2.6  $\mu\text{mol}$ , 1.5 equiv.) were added. Degassed deionized  $\text{H}_2\text{O}$  was added up to 500  $\mu\text{L}$  total volume. The reaction was run under nitrogen atmosphere and monitored by analytical HPLC and LCMS. After 4 h the reaction was considered complete and solvents were removed by a nitrogen stream. The remaining solid was diluted with 500  $\mu\text{L}$  of buffer A and the solution was centrifuged (4500 rpm, 5 min) and purified using a gradient of 0-40% of buffer B into buffer A over 40 min at a flow rate of 12.5  $\text{mL}\cdot\text{min}^{-1}$ . The final compound **22** was obtained as a white fluffy solid after lyophilization (3.2 mg, 0.61  $\mu\text{mol}$ , 35%, overall yield: 9%). aHPLC  $t_{\text{R}}$  = 16.434, LCMS-ESI:  $t_{\text{R}}$  = 11.64, average mass calculated  $[\text{M}+4\text{H}]^{4+}$ : 1191.84; found: 1191.67. Average mass calculated  $[\text{M}+\text{H}]^+$ : 4764.33, MALDI-TOF found: 4764.28.

### Synthesis gp 120 mimic 30:



*First CuAAC, TES-removal and second CuAAC were conducted in one pot as follows:*

*Step 1, first CuAAC: TAC-scaffold **15** (7.88 mg, 10  $\mu\text{mol}$ , 1 equiv.), peptide loop **3** **18** (14.7 mg, 10.7  $\mu\text{mol}$ , 1.07 equiv.) and TBTA (2.7 mg, 5.0  $\mu\text{mol}$ , 0.5 equiv.)*



were placed in a flask and degassed DMF (300  $\mu\text{L}$ ) was added. The reaction was stirred under nitrogen atmosphere until all components were dissolved. Then, aqueous 1 M  $\text{CuSO}_4$  (5  $\mu\text{L}$ , 5.0  $\mu\text{mol}$ , 0.5 equiv.) and aqueous 0.25 M sodium ascorbate (60  $\mu\text{L}$ , 15 mmol, 1.5 equiv.) were added. Degassed deionized  $\text{H}_2\text{O}$  was added up to 500  $\mu\text{L}$  of volume to the reaction mixture. The reaction mixture was stirred under nitrogen and monitored by analytical HPLC and LCMS. When reaction was complete (typically after 3.5 h) the subsequent deprotection step was started.

*Step 2, TES removal:*  $\text{AgNO}_3$  (20.4 mg, 120  $\mu\text{mol}$ , 12 equiv.) was added to the above mixture and stirred under a nitrogen atmosphere. Reaction progress was followed by analytical HPLC and LCMS, which showed complete conversion after 1 h. Then,  $\text{NaCl}$  (7 mg, 120  $\mu\text{mol}$ , 12 equiv.) was added in order to precipitate out  $\text{Ag}^+$  as  $\text{AgCl}$ . This suspension was used for the subsequent (second) CuAAC step.

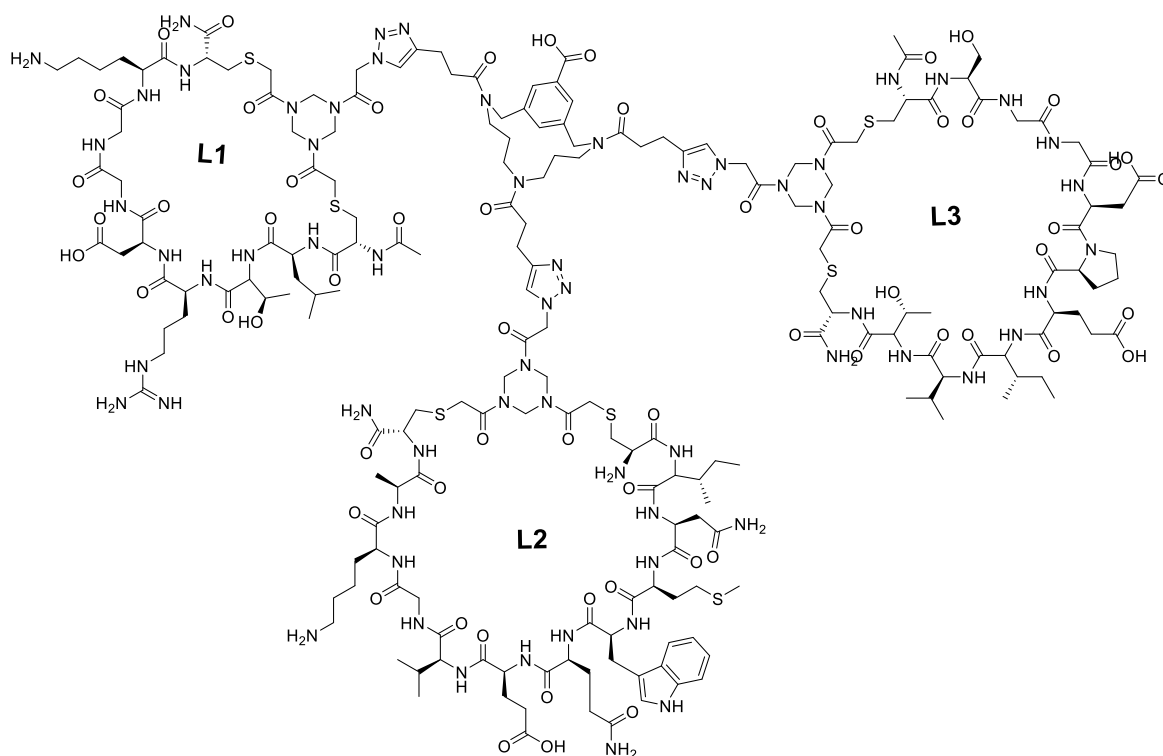
*Step 3, second CuAAC:* To the above suspension, peptide loop 1 **16** (16.2 mg, 10.7  $\mu\text{mol}$ , 1.07 equiv.), TBTA (2.7 mg, 5.0  $\mu\text{mol}$ , 0.5 equiv.), aqueous 1 M  $\text{CuSO}_4$  (5  $\mu\text{L}$ , 5.0  $\mu\text{mol}$ , 0.5 equiv.) and aqueous 0.25 M sodium ascorbate (60  $\mu\text{L}$ , 15 mmol, 1.5 equiv.) were added. The resulting mixture was stirred under nitrogen atmosphere and monitored by analytical HPLC and LCMS. The reaction was stirred overnight, and solvents removed by a nitrogen stream. A red liquid residue was diluted with buffer A and the resulting solution was centrifuged (4500 rpm, 5 min) and purified using a gradient of 20-70% of buffer B into buffer A over 50 min at 12.5  $\text{mL}\cdot\text{min}^{-1}$ . The product containing fractions were pooled and lyophilized to obtain the scaffold derivative **28** with two peptides attached as a white fluffy solid (12.4 mg, 3.53  $\mu\text{mol}$ , 71% per step, 36% over 3 steps). aHPLC  $t_R$  = 21.484, LCMS-ESI:  $t_R$  = 16.66, average mass calculated  $[\text{M}+3\text{H}]^{3+}$ : 1097.27; found: 1097.00.

*TIPS-removal:* TIPS-protected scaffold **28**  $\times$  2-TFA (12.4 mg, 3.53  $\mu\text{mol}$ ) and TBAF $\cdot$ 3 $\text{H}_2\text{O}$  (44.55 mg, 141.2  $\mu\text{mol}$ , 40 equiv.) were weighed out into a vial and dissolved in degassed DMF (1491  $\mu\text{L}$ ). The reaction mixture was stirred under nitrogen atmosphere and monitored by analytical HPLC and LCMS. After 2.5 h, when the reaction was complete, DMF was removed with nitrogen stream. The remaining solid was diluted with buffer A (520  $\mu\text{L}$ ) and centrifuged (4500 rpm, 5 min). The resulting solution was purified using a gradient of 0-40% of buffer B

into buffer A over 40 min at  $12.5 \text{ mL} \cdot \text{min}^{-1}$ . TIPS-deprotected product was obtained as a white fluffy solid of product **29** (7.4 mg,  $2.20 \text{ } \mu\text{mol}$ , 62%) after lyophilization. aHPLC  $t_R = 15.616$ , LCMS-ESI:  $t_R = 11.12$  average mass calculated  $[\text{M}+3\text{H}]^{3+}$ : 1044.82; found: 1044.92.

*Third CuAAC:* TIPS-deprotected scaffold **29** (8.4 mg,  $2.50 \text{ } \mu\text{mol}$ , 1 equiv.), peptide loop 2 **17** (4.88 mg,  $2.63 \text{ } \mu\text{mol}$ , 1.05 equiv.), TBTA (1.32 mg,  $2.5 \text{ } \mu\text{mol}$ , 1 equiv.) were placed in a vial and dissolved in degassed DMF ( $300 \text{ } \mu\text{L}$ ) and stirred under nitrogen atmosphere until dissolved. Then, aqueous 1 M  $\text{CuSO}_4$  ( $2.5 \text{ } \mu\text{L}$ ,  $2.5 \text{ } \mu\text{mol}$ , 1 equiv.) and aqueous 0.25 M sodium ascorbate ( $28.5 \text{ } \mu\text{L}$ ,  $3.75 \text{ } \mu\text{mol}$ , 1.5 equiv.) were added. Degassed deionized  $\text{H}_2\text{O}$  was added up to  $500 \text{ } \mu\text{L}$  of volume of the reaction mixture. The reaction was run under nitrogen atmosphere and monitored by analytical HPLC and LCMS. The reaction was stirred for 19 h and solvents were removed by a nitrogen stream. The remaining solid was diluted with  $300 \text{ } \mu\text{L}$  of buffer A and the solution was centrifuged (4500 rpm, 5 min) and purified using a gradient of 18-38% of buffer B into buffer A over 40 min at a flow rate of  $10.0 \text{ mL} \cdot \text{min}^{-1}$ . The final gp120 mimic TAC-scaffold containing 3 loops **30** was obtained as a white fluffy solid after lyophilization (1.2 mg,  $0.23 \text{ } \mu\text{mol}$ , 9%). aHPLC  $t_R = 16.436$ , LCMS-ESI:  $t_R = 11.65$ , average mass calculated  $[\text{M}+4\text{H}]^{4+}$ : 1191.84; found: 1191.67. Average mass calculated  $[\text{M}+\text{H}]^+$ : 4764.3292, MALDI-TOF found: 4764.82.

## Synthesis gp120 mimic 26:



*First cycloaddition, TES-removal and second cycloaddition in one pot:*

**Step 1, First CuAAC:** TAC-scaffold **15** (4.83 mg, 6.13  $\mu\text{mol}$ , 1 equiv.), peptide loop **2** **17** (11.4 mg, 6.13  $\mu\text{mol}$ , 1 equiv.) and TBTA (1.64 mg, 3.0  $\mu\text{mol}$ , 0.5 equiv.) were placed in a flask and degassed DMF (300  $\mu\text{L}$ ) was added. The reaction was stirred under nitrogen atmosphere until all components were dissolved. Then, aqueous 1 M  $\text{CuSO}_4$  (3.1  $\mu\text{L}$ , 3.0  $\mu\text{mol}$ , 0.5 equiv.) and aqueous 0.25 M sodium ascorbate (36.4  $\mu\text{L}$ , 9.2 mmol, 1.5 equiv.). The reaction mixture was stirred under nitrogen and monitored by analytical HPLC and LCMS. When reaction was complete (typically after 3 h) the subsequent deprotection step was started.

**Step 2, TES removal:**  $\text{AgNO}_3$  (12.5 mg, 73.6  $\mu\text{mol}$ , 12 equiv.) was added and the mixture was stirred under nitrogen atmosphere, and monitored by analytical HPLC and LCMS, which showed complete conversion after 1 h. Then,  $\text{NaCl}$  (4 mg, 73.6  $\mu\text{mol}$ , 12 equiv.) was added in order to precipitate out  $\text{Ag}^+$  as  $\text{AgCl}$ . This suspension was used for the subsequent second CuAAC step.

**Step 3, Second CuAAC:** Peptide loop **1** **16** (9.0 mg, 6.13  $\mu\text{mol}$ , 1 equiv.), TBTA (3.28 mg, 6.13  $\mu\text{mol}$ , 1 equiv.), aqueous 1 M  $\text{CuSO}_4$  (6.1  $\mu\text{L}$ , 6.13  $\mu\text{mol}$ , 1 equiv.)

and aqueous 0.25 M sodium ascorbate (72.8  $\mu\text{L}$ , 18.4  $\mu\text{mol}$ , 3 equiv.) were added and the resulting mixture was stirred under nitrogen atmosphere and monitored by analytical HPLC and LCMS. After 2 h the reaction was complete, and solvents removed by a nitrogen stream. A red liquid residue was diluted with buffer A and the resulting solution was centrifuged (4500 rpm, 5 min) and purified using a gradient of 30-60% of buffer B into buffer A over 40 min at 12.5  $\text{mL}\cdot\text{min}^{-1}$ . The product containing fractions were pooled and lyophilized to obtain the scaffold derivative **24** with two peptides attached as a white fluffy solid (5.4 mg, 1.35  $\mu\text{mol}$ , 60% per step, 22% over 3 steps). aHPLC  $t_{\text{R}}$  = 21.165 (in crude reaction mixture), LCMS-ESI:  $t_{\text{R}}$  = 16.26, average mass calculated  $[\text{M}+3\text{H}]^{3+}$ : 1183.75; found: 1183.58.

*Step 3, TIPS removal:* TIPS-protected scaffold **24**  $\times$  4·TFA (5.4 mg, 1.34  $\mu\text{mol}$ ) and TBAF·3H<sub>2</sub>O (17 mg, 54  $\mu\text{mol}$ , 40 equiv.) were weighed out into a vial and dissolved in degassed DMF (570  $\mu\text{L}$ ). The reaction mixture was stirred under nitrogen atmosphere and monitored by analytical HPLC and LCMS. After 4 h, when reaction was complete, DMF was removed with nitrogen stream. The remaining solid was diluted with buffer A (300  $\mu\text{L}$ ) and centrifuged (4500 rpm, 5 min). The resulting solution was purified using a gradient of 0-50% of buffer B into buffer A over 50 min at 12.5  $\text{mL}\cdot\text{min}^{-1}$ . TIPS-deprotected product **25** was obtained as a white fluffy solid (3.6 mg, 0.94  $\mu\text{mol}$ , 69%) after lyophilization. aHPLC  $t_{\text{R}}$  = 16.666, LCMS-ESI:  $t_{\text{R}}$  = 11.59, average mass calculated  $[\text{M}+3\text{H}]^{3+}$ : 1131.23; found: 1131.50.

*Step 4, Third CuAAC:* TIPS-deprotected scaffold **25** (3.6 mg, 0.94  $\mu\text{mol}$ , 1 equiv.), peptide loop 3 **18** (1.3 mg, 0.94  $\mu\text{mol}$ , 1 equiv.), TBTA (0.5 mg, 0.94  $\mu\text{mol}$ , 1 equiv.) were placed in a vial and dissolved in degassed DMF (300  $\mu\text{L}$ ) and stirred under nitrogen atmosphere until dissolved. Then, aqueous 1 M CuSO<sub>4</sub> (1  $\mu\text{L}$ , 1.1  $\mu\text{mol}$ , 1.1 equiv.) and aqueous 0.25 M sodium ascorbate (16.6  $\mu\text{L}$ , 4.2  $\mu\text{mol}$ , 4.5 equiv.) were added. Degassed deionized H<sub>2</sub>O was added up to 500  $\mu\text{L}$  of volume of the reaction mixture. The reaction was run under nitrogen atmosphere and monitored by analytical HPLC and LCMS. After 2.5 h the reaction was considered complete and solvents were removed by a nitrogen stream. The remaining solid was diluted with 300  $\mu\text{L}$  of buffer A and the solution was centrifuged (4500 rpm, 5 min) and purified using a gradient of 0-40% of buffer B into buffer A over 40 min at a flow rate of 12.5  $\text{mL}\cdot\text{min}^{-1}$ . The final gp120 mimic TAC-scaffold containing 3

loops **26** was obtained as a white fluffy solid after lyophilization (3.4 mg, 0.65  $\mu\text{mol}$ , 69%). aHPLC  $t_R$  = 16.423, LCMS-ESI:  $t_R$  = 11.62, average mass calculated  $[\text{M}+4\text{H}]^{4+}$ : 1191.84; found: 1191.67. Average mass calculated  $[\text{M}+\text{H}]^+$ : 4764.33, MALDI-TOF found: 4764.82.

**Synthesis gp120 mimic **26** by reversed order of attachment of loop 1 and 3:**

*First cycloaddition, TES-removal and second cycloaddition in one pot:* TAC-scaffold **15** (6.61 mg, 8.4  $\mu\text{mol}$ , 1 equiv.equiv.), peptide loop 2 **17** (15.6 mg, 8.4  $\mu\text{mol}$ , 1 equiv. and TBTA (1.0 mg, 4.2  $\mu\text{mol}$ , 0.5 equiv.) were placed in a flask and degassed DMF (300  $\mu\text{L}$ ) was added. The reaction was stirred under nitrogen atmosphere until all components were dissolved. Then, aqueous 1 M  $\text{CuSO}_4$  (4.2  $\mu\text{L}$ , 4.2  $\mu\text{mol}$ , 0.5 equiv.) and aqueous 0.25 M sodium ascorbate (50  $\mu\text{L}$ , 12.6 mmol, 1.5 equiv.). The reaction mixture was stirred under nitrogen and monitored by analytical HPLC and LCMS. When reaction was complete (typically after 2.5 h),  $\text{AgNO}_3$  (17.1 mg, 100.8  $\mu\text{mol}$ , 12 equiv.) was added and the mixture was stirred under nitrogen atmosphere, and monitored by analytical HPLC and LCMS, which showed complete conversion after 1h. Then,  $\text{NaCl}$  (5.9 mg, 100.8  $\mu\text{mol}$ , 12 equiv.) was added in order to precipitate out  $\text{Ag}^+$  as  $\text{AgCl}$ . Afterwards, peptide loop 3 **18** (11.5 mg, 8.4  $\mu\text{mol}$ , 1 equiv.), TBTA (2.2 mg, 4.2  $\mu\text{mol}$ , 1 equiv.), aqueous 1 M  $\text{CuSO}_4$  (4.2  $\mu\text{L}$ , 4.2  $\mu\text{mol}$ , 1 equiv.) and aqueous 0.25 M sodium ascorbate (50  $\mu\text{L}$ , 12.6  $\mu\text{mol}$ , 1.5 equiv.) were added and the resulting mixture was stirred under nitrogen atmosphere and monitored by analytical HPLC and LCMS. After 4h 0.25 M sodium ascorbate was added (50  $\mu\text{L}$ , 12.6  $\mu\text{mol}$ , 1.5 equiv.). When reaction was not finished after next 5 h, another portion of 1 M  $\text{CuSO}_4$  (4.2  $\mu\text{L}$ , 4.2  $\mu\text{mol}$ , 0.5 equiv.), TBTA (2.2 mg, 4.2  $\mu\text{mol}$ , 0.5 equiv.) and 0.25 sodium ascorbate (125  $\mu\text{L}$ , 31.5  $\mu\text{mol}$ , 3.8 equiv.) were added. The reaction was stirred further under nitrogen atmosphere, and after 2 h another portion of 0.25 M sodium ascorbate (50  $\mu\text{L}$ , 12.6  $\mu\text{mol}$ , 1.5 equiv.) was added. Then, after 1 h the reaction was complete and solvents removed by a nitrogen stream. A red liquid residue was diluted with buffer A (700  $\mu\text{L}$ ) and the resulting solution was centrifuged (4500 rpm, 5 min) and purified using a gradient of 10-60% of buffer B into buffer A over 50 min at a flow rate of 12.5 ml min<sup>-1</sup>. The product containing fractions were pooled and lyophilized to obtain the scaffold derivative **31** with two peptides attached as a white fluffy solid (3.65 mg, 0.94  $\mu\text{mol}$ , 48% per step, 11.2% over 3

steps).  $t_R = 22.036$ , LCMS-ESI: average mass calculated  $[M+3H]^{3+}$ : 1225.5406; found: 1226.17.

*Removal of the TIPS-protecting group:* TIPS-protected scaffold **31**  $\times$  4TFA (3.65 mg, 0.94  $\mu$ mol) and TBAF $\cdot$ 3H<sub>2</sub>O (11.9 mg, 37.6  $\mu$ mol, 40 equiv.) were weighed out into a vial and dissolved in degassed DMF (125  $\mu$ L, to obtain concentration of 95 mM of TBAF $\cdot$ 3H<sub>2</sub>O in DMF). The reaction mixture was stirred under nitrogen atmosphere and monitored by analytical HPLC and LCMS. After 4 h, when reaction was complete, DMF was removed with nitrogen stream. The remaining solid was diluted with buffer A (160  $\mu$ L) and centrifuged (4500 rpm, 5 min). The resulting solution was purified using a gradient of 0-40% over 40 min at a flow rate of 12.5 ml min<sup>-1</sup>. TIPS-deprotected **32** product was obtained as a white fluffy solid (1.9 mg, 0.51  $\mu$ mol, 54%) after lyophilization.  $t_R = 17.319$ , LCMS-ESI: average mass calculated  $[M+3H]^{3+}$ : 1173.9883; found: 1174.00.

*Cycloaddition of the third peptide onto the scaffold:* TIPS-deprotected scaffold **32** (1.9 mg, 0.51  $\mu$ mol, 1 equiv.), peptide loop 1 **16** (0.8 mg, 0.51  $\mu$ mol, 1 equiv.), TBTA (0.9 mg, 1.5  $\mu$ mol, 3 equiv.) were placed in a vial and dissolved in degassed DMF (300  $\mu$ L) and stirred under nitrogen atmosphere until dissolved. Then, aqueous 1 M CuSO<sub>4</sub> (1.5  $\mu$ L, 1.5  $\mu$ mol, 3.0 equiv.) and aqueous 0.25 M sodium ascorbate (18.0  $\mu$ L, 4.6  $\mu$ mol, 9.0 equiv.) were added. Degassed deionized H<sub>2</sub>O was added up to 500  $\mu$ L of volume of the reaction mixture. The reaction was run under nitrogen atmosphere and monitored by analytical HPLC and LCMS. After 4 h the reaction was considered complete and solvents were removed by a nitrogen stream. The remaining solid was diluted with 400  $\mu$ L of buffer A and the solution was centrifuged (4500 rpm, 5 min) and purified using a gradient of 0-40% of buffer B into buffer A over 40 min at a flow rate of 12.5 mL min<sup>-1</sup>. The final gp120 mimic TAC-scaffold containing 3 loops **26** was obtained as a white fluffy solid after lyophilization (0.6 mg, 0.12  $\mu$ mol, 22.5%).  $t_R = 16.381$ , LCMS-ESI: average mass calculated  $[M+4H]^{4+}$ : 1191.8395; found: 1191.75.

# Appendix

Accession	Description	Score	Coverage	# Proteins	# Unique Peptides
P01730	T-cell surface glycoprotein CD4 OS=Homo sapiens GN=CD4 PE=1 SV=1 - [CD4_HUMAN]	111.19	10.70	1	5
A3					
	High	NSNQIKILGNQGSFLTK	1	1	1
	High	KGDTVELTCTASQK	2	1	1
	High	IDIVLAFQK	11	1	1
	High	ILGNQGSFLTK	24	1	1
	High	KSIQFHWK	1	1	1

# Peptides	# PSMs	Area	# AAs	MW [kDa]	
5	39	7.386E8	458	51.1	
Protein Group Accessions					
P01730		0.0000	2.023E6	5.78	
P01730		0.0000	2.266E6	4.97	
P01730		0.0000	2.324E8	3.40	
P01730		0.0000	1.981E9	3.30	
P01730		0.0000	2.665E6	2.70	
calc. pI					
	9.54				
Probability					
	0.00	3	1862.00846	-2.50	45.10
	0.00	3	1480.72922	-1.37	28.84
	0.00	2	1145.68999	-2.57	56.34
	0.00	2	1177.65544	-1.83	43.00
	0.00	3	1073.58899	-0.11	31.43
# Missed Cleavages					
					1
					1
					0
					0
					1

Figure 97 Results of protein mass spectrometry of CD4D12.

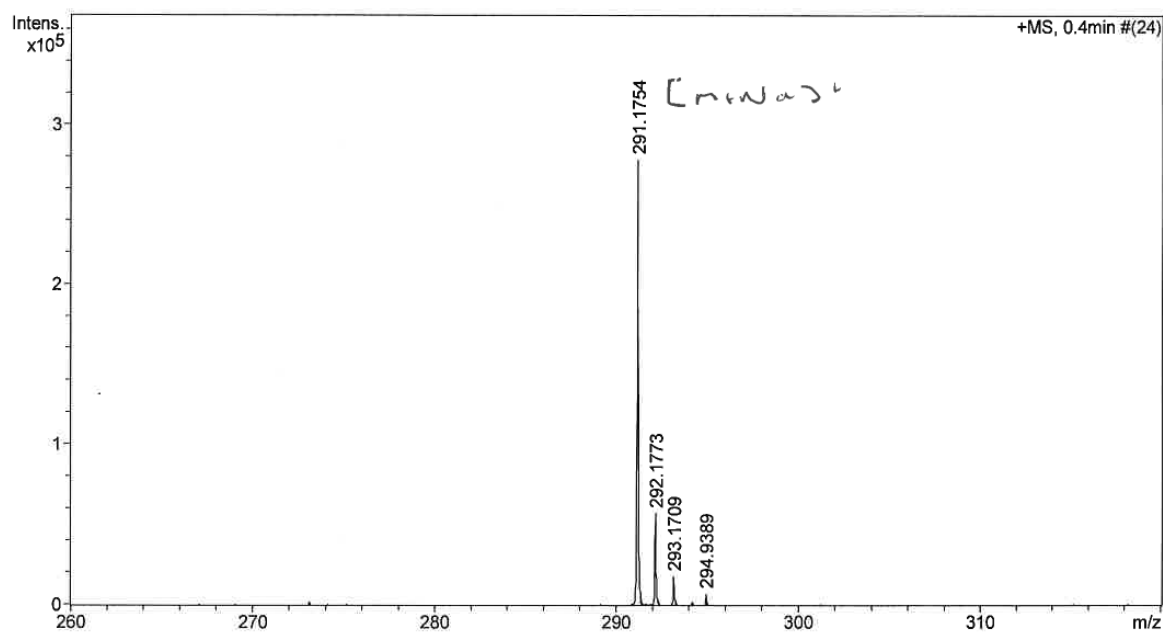


Figure 98 HRMS of Tert-butyl 5-(triethylsilyl)-pent-4-ynoate.

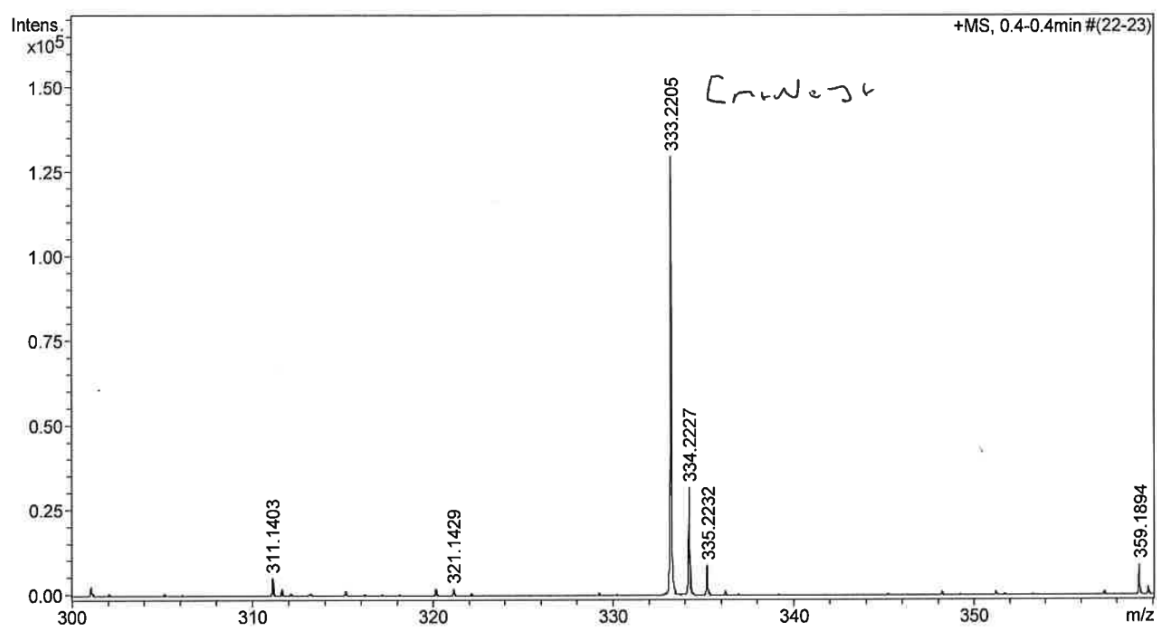


Figure 99 HRMS of Tert-butyl 5-(triisopropylsilyl)-pent-4-ynoate.



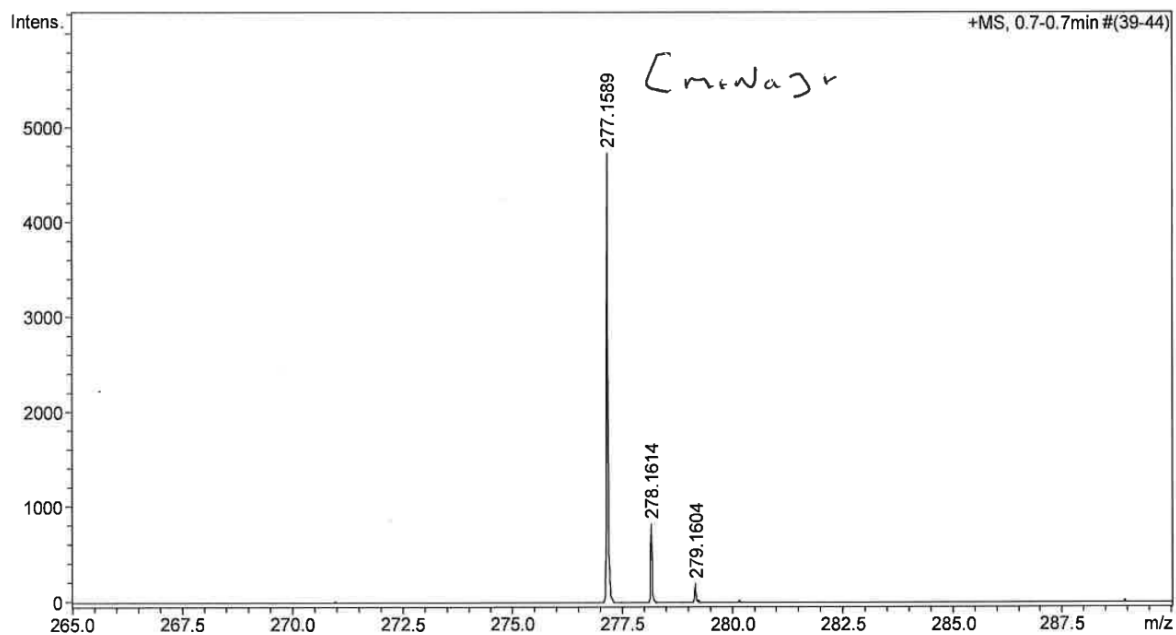


Figure 100 HRMS of 5-(triisopropylsilyl)-4-pentynoic acid.

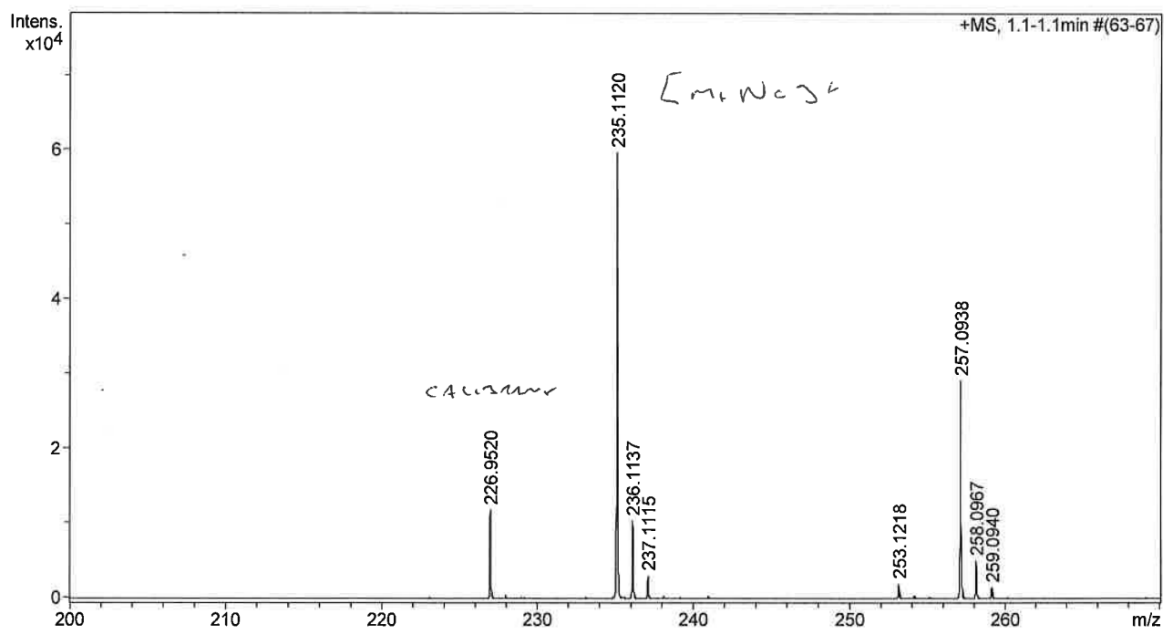
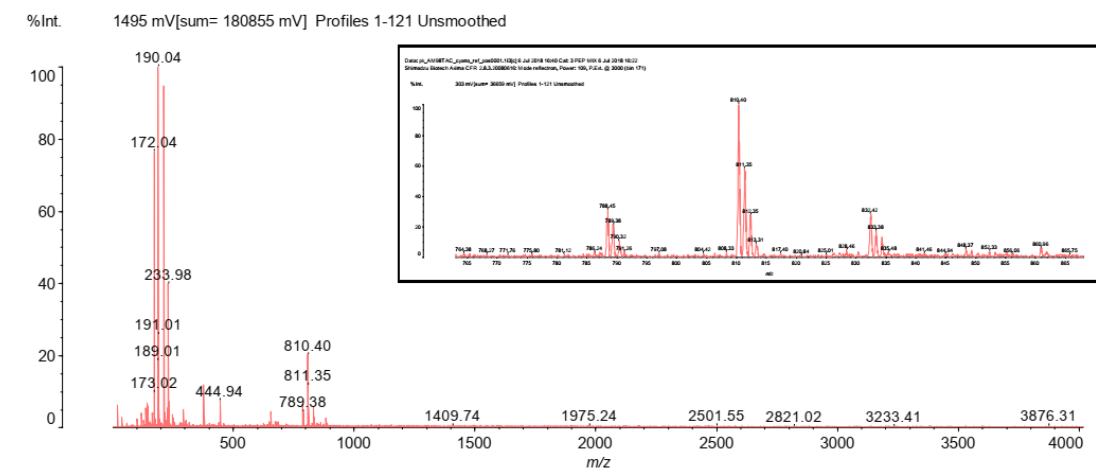


Figure 101 HRMS of 5-(triethylsilyl)-4-pentynoic acid.

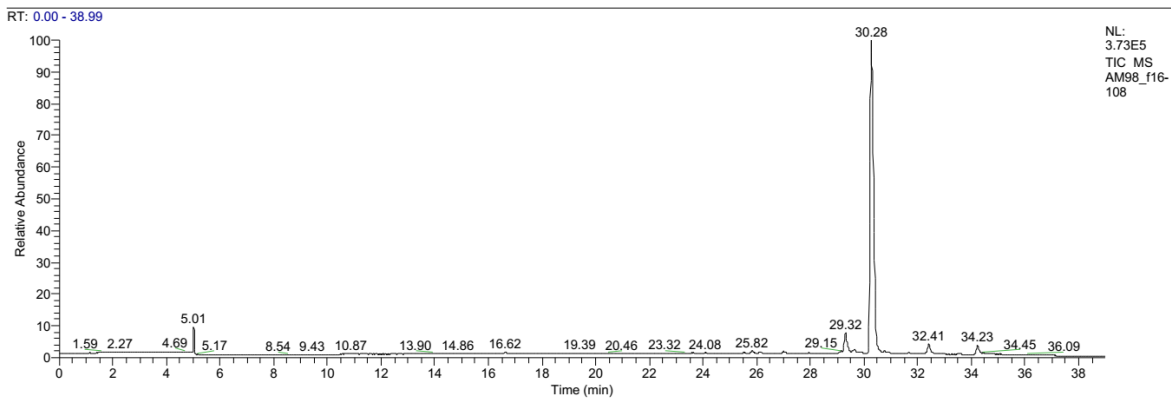


Mass spectrum of compound 10b at 1.5 minutes. The x-axis is  $m/z$  from 750 to 840, and the y-axis is Intensity ( $\times 10^4$ ) from 0.0 to 1.5. The base peak is at  $m/z$  810.4628, labeled  $[M+Na]^+$ . Other significant peaks are at  $m/z$  770.8525, 788.4784, 801.6589, 804.8464, 819.9995, 832.4432, and 838.8425.

Figure 103 HRMS of orthogonally protected TAC scaffold **15**.

E:\LCMS\...\AM98\_f16-108

10/04/2018 11:30:00



AM98\_f16-108 #1795 RT: 30.29 AV: 1 NL: 1.28E4  
T: ITMS + p ESI Full ms [150.00-2000.00]

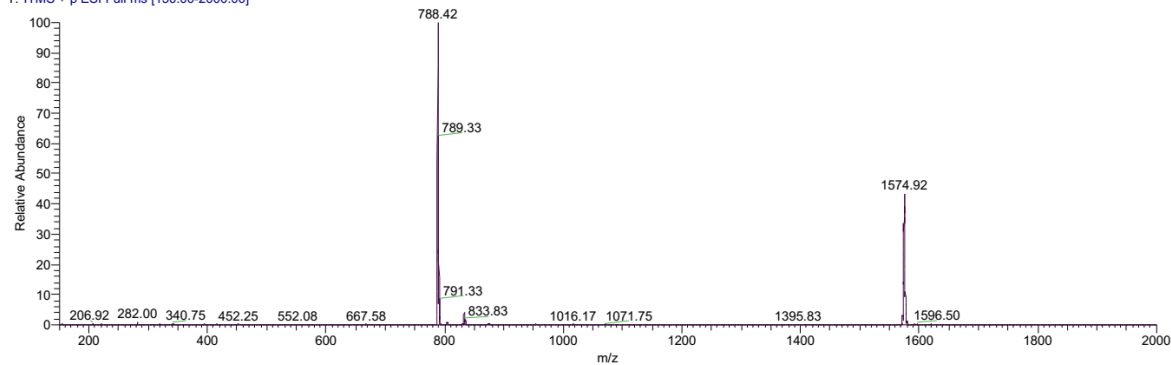


Figure 104 LCMS of orthogonally protected TAC scaffold 15.

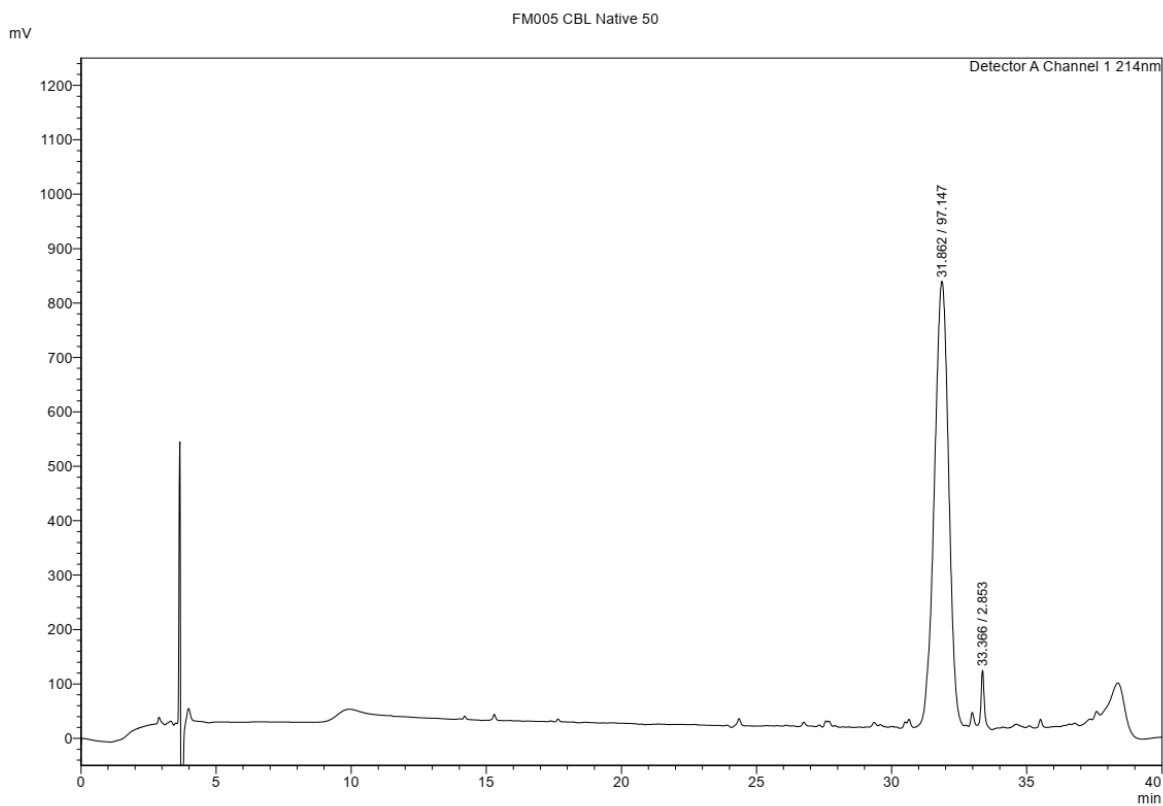


Figure 105 analytical HPLC of orthogonally protected TAC scaffold 15.

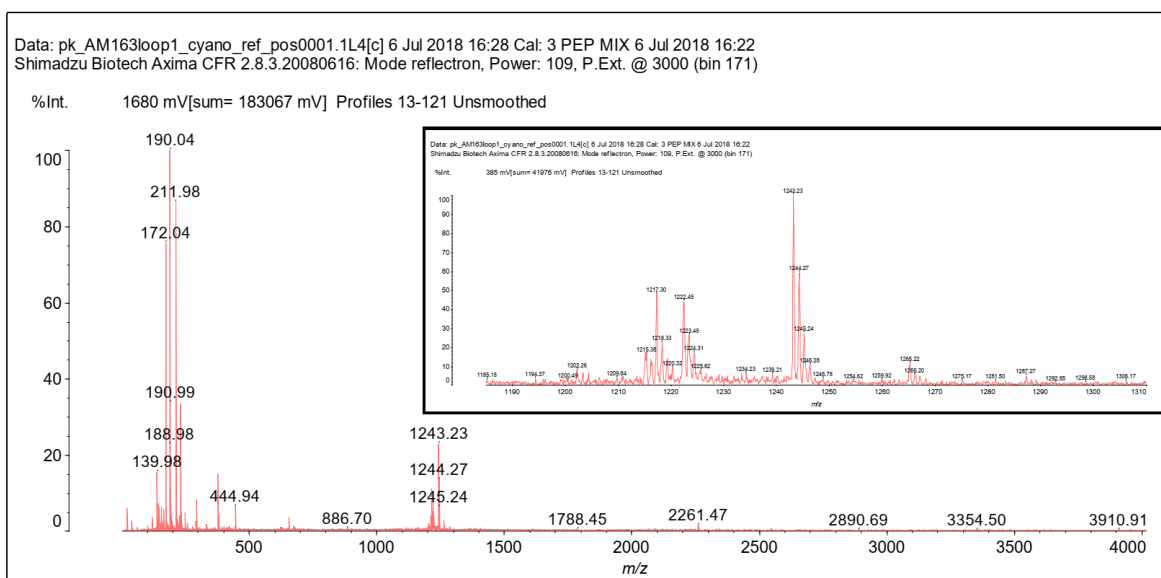


Figure 106 MALDI-TOF of loop 1.

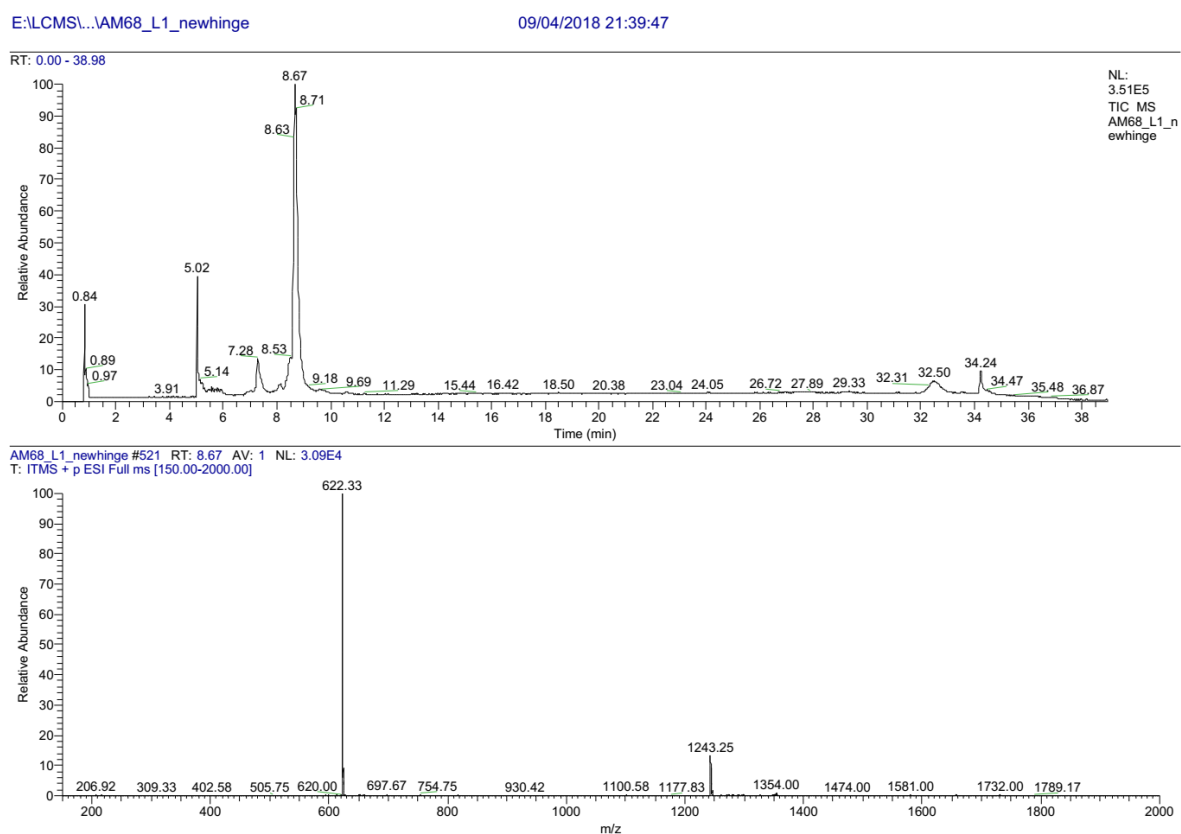


Figure 107 LCMS of loop 1.

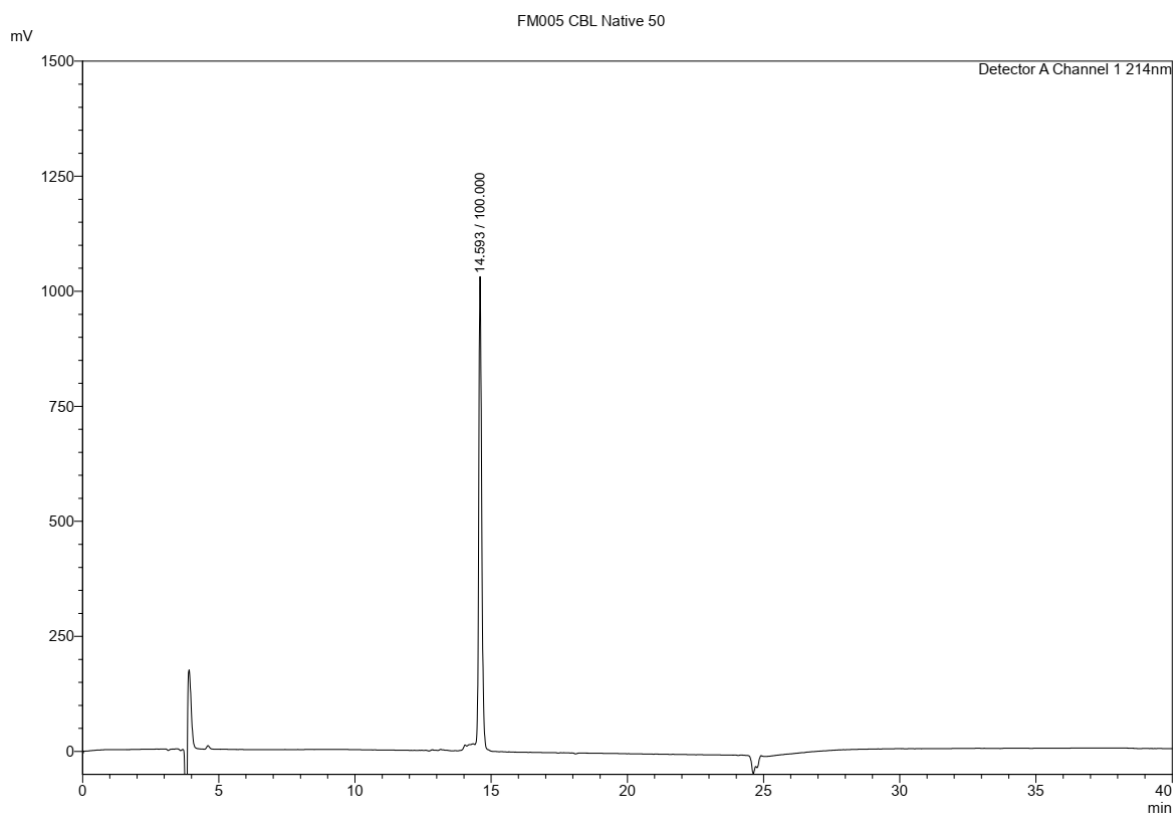


Figure 108 analytical HPLC of loop 1.

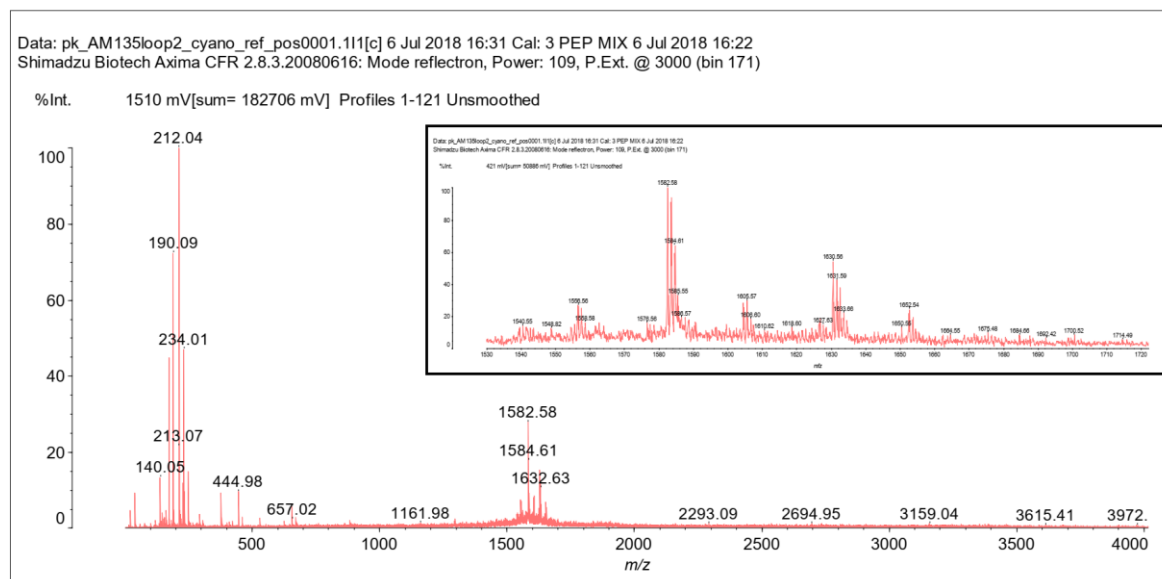


Figure 109 MALDI-TOF of loop 2.

E:\LCMS\...\peptides\_new\AM144\_pureR1-R7

01/06/2018 18:04:58

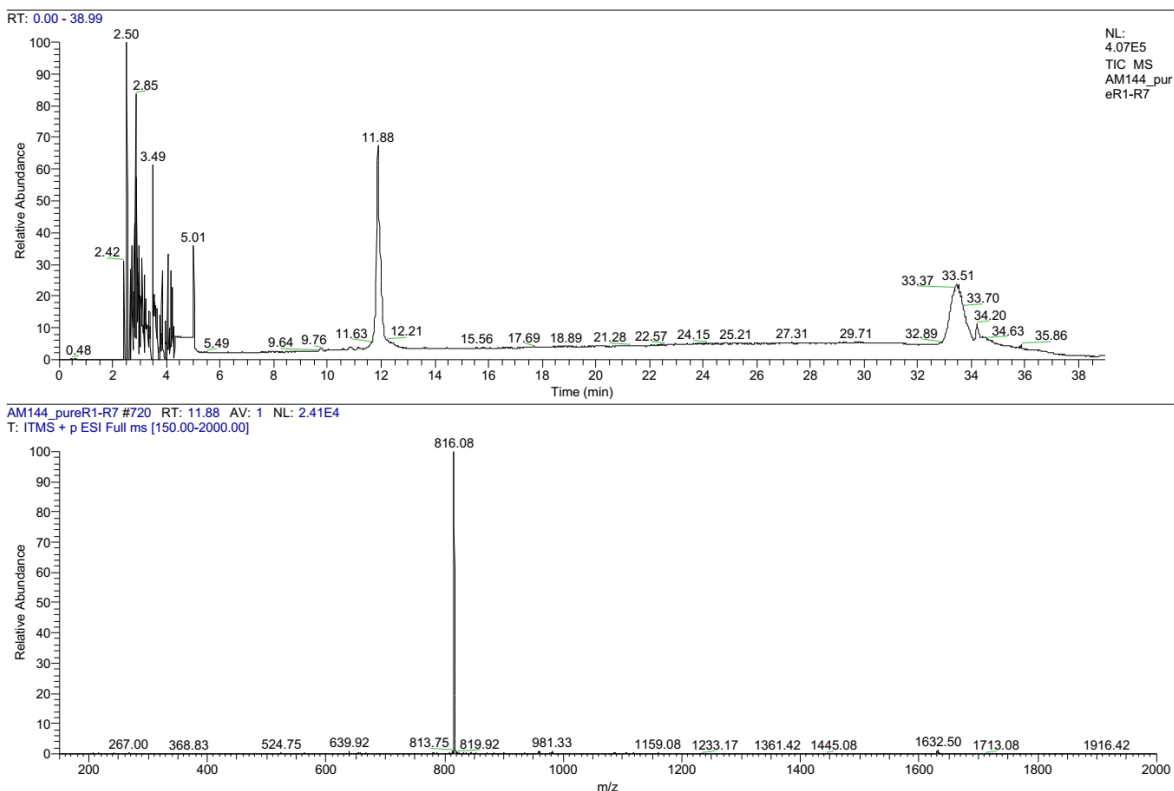


Figure 110 LCMS of loop 2.

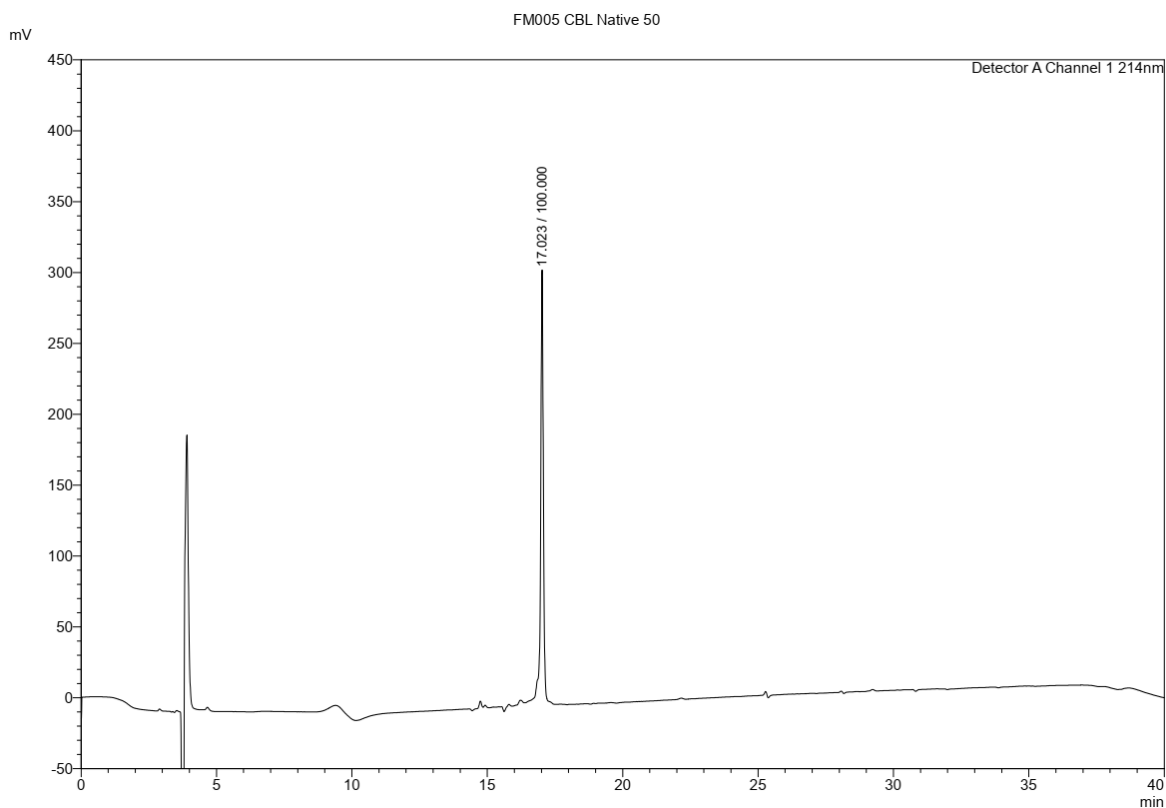


Figure 111 analytical HPLC of loop 2.

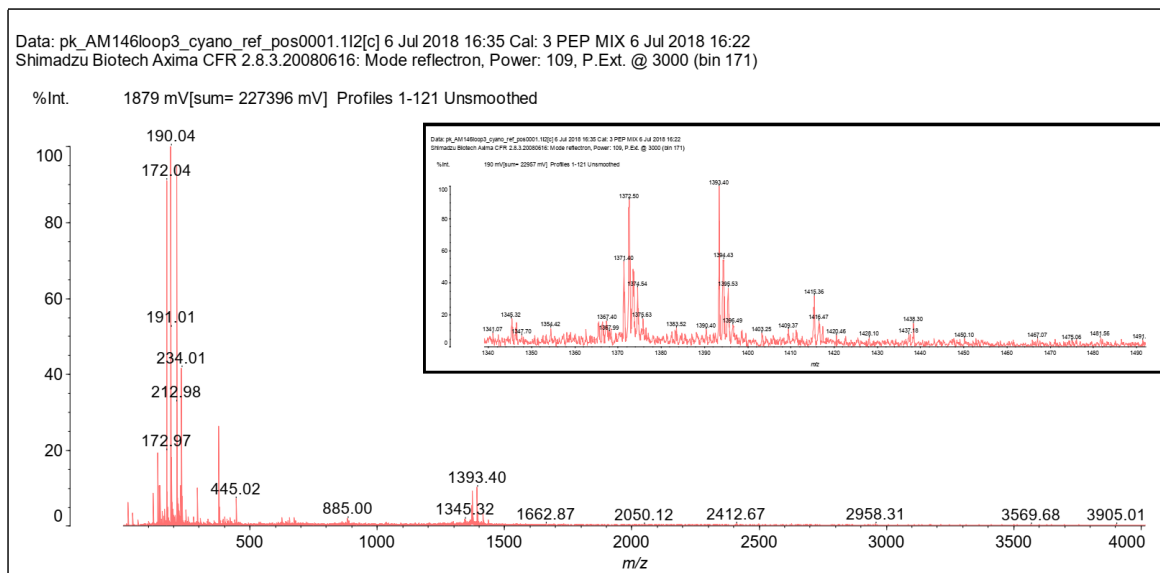


Figure 112 MALDI-TOF of loop 3.

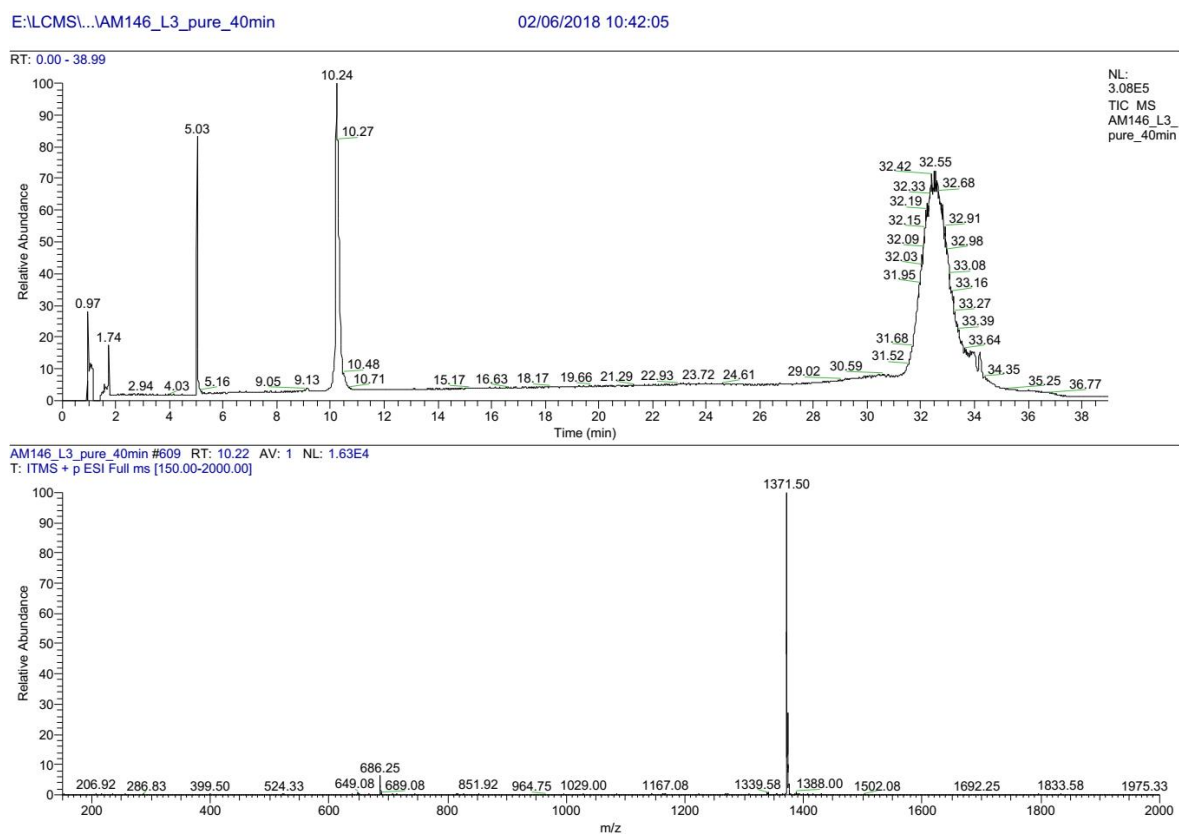


Figure 113 LCMS of loop 3.

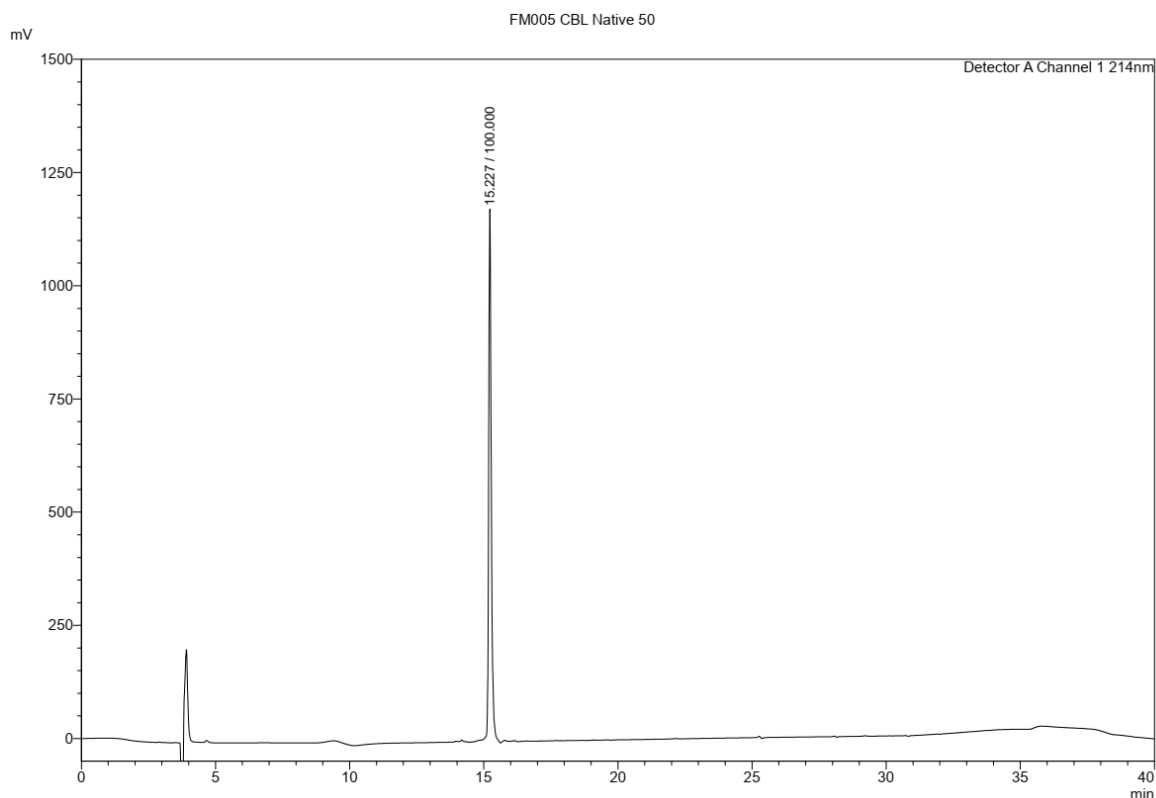


Figure 114 analytical HPLC of loop 3.

E:\LCMS\...\AM102\_crude\_3h

12/04/2018 13:51:22

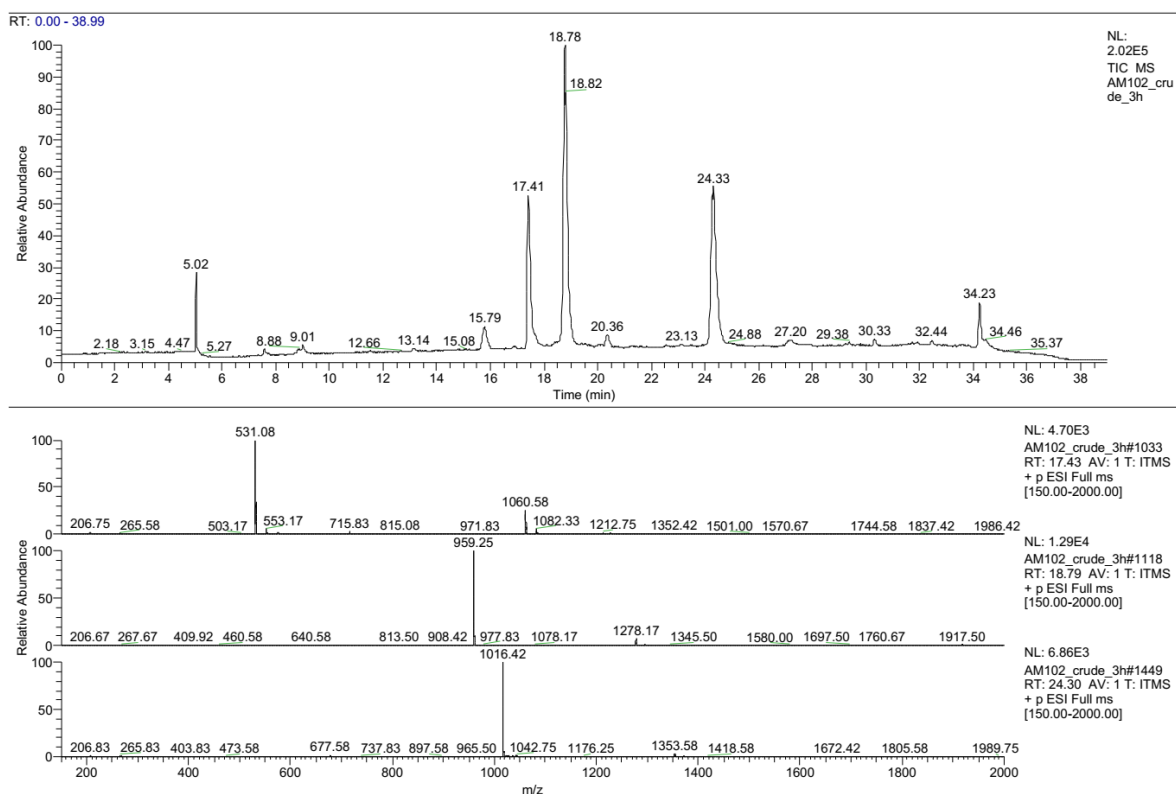


Figure 115 LCMS of cycloaddition of the first loop in the synthesis of construct 22.



E:\LCMS\...AM103\_f93-100\_pure

16/04/2018 13:47:30

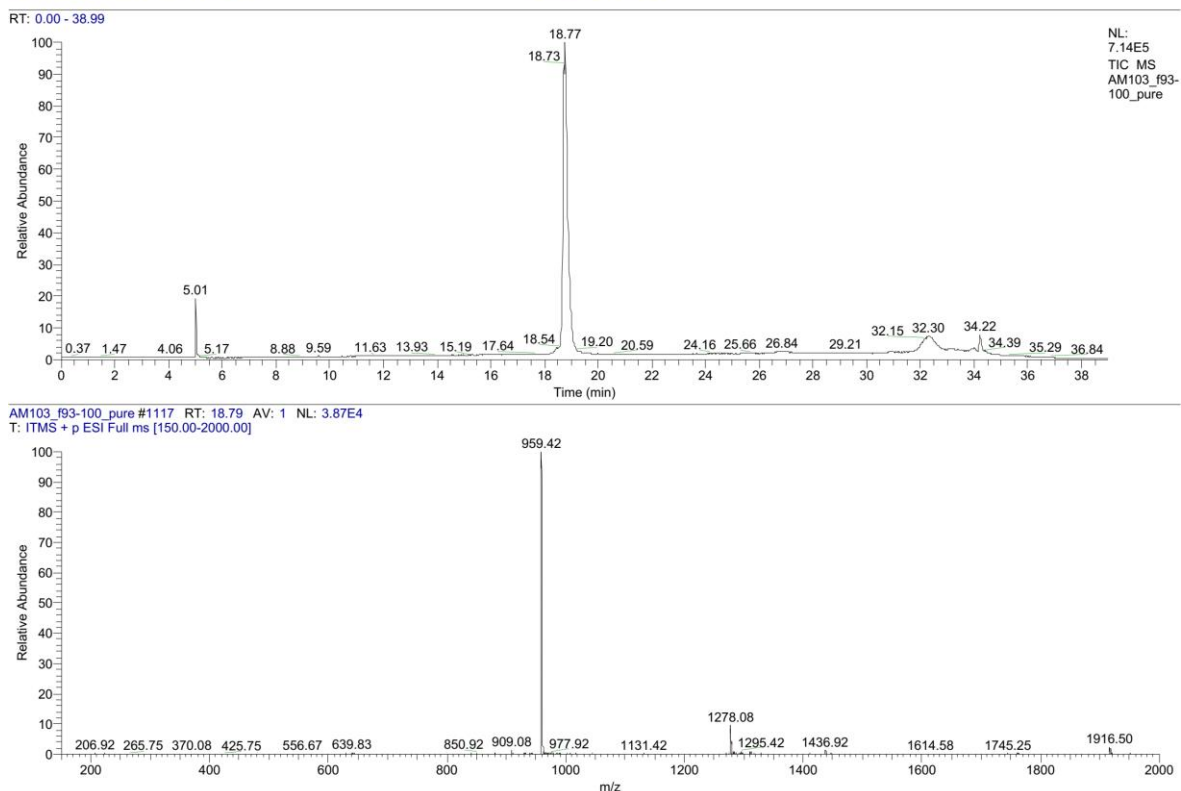


Figure 116 LCMS of TES removal to obtain product 19.

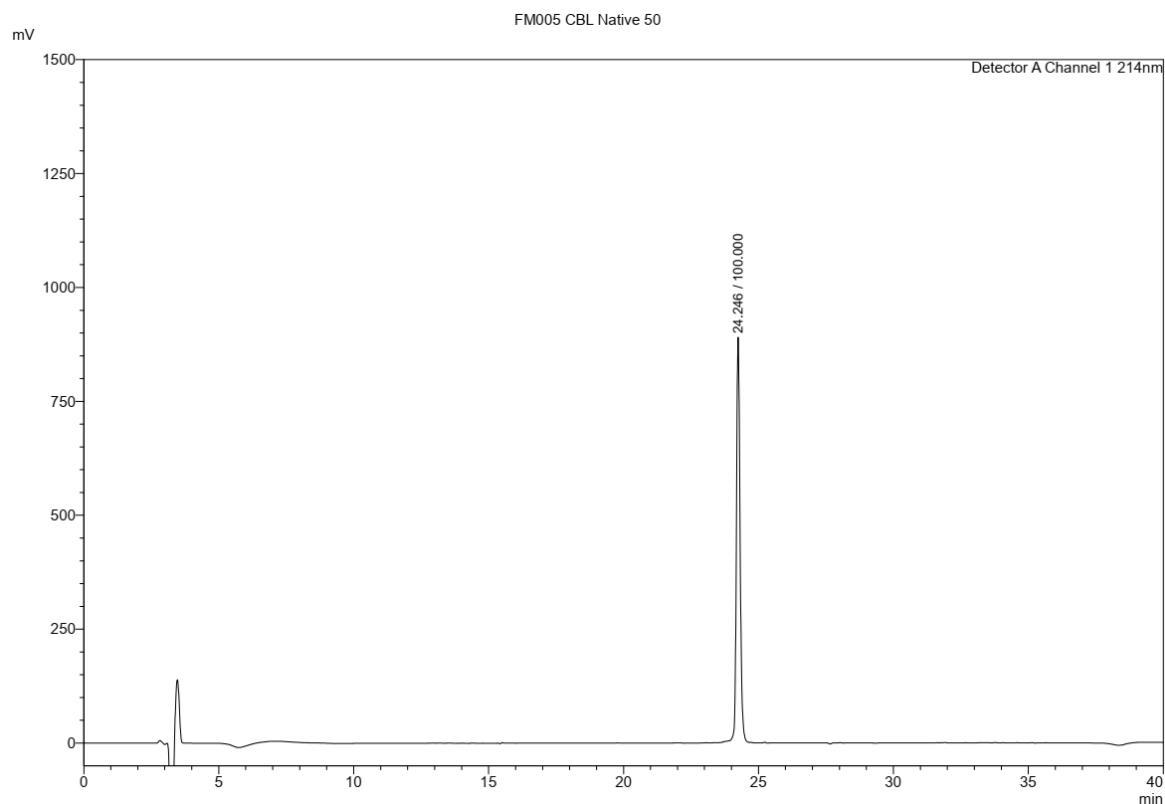


Figure 117 analytical HPLC of purified compound 19.

E:\LCMS\...\AM105\_f50-64\_purified

23/04/2018 13:11:13

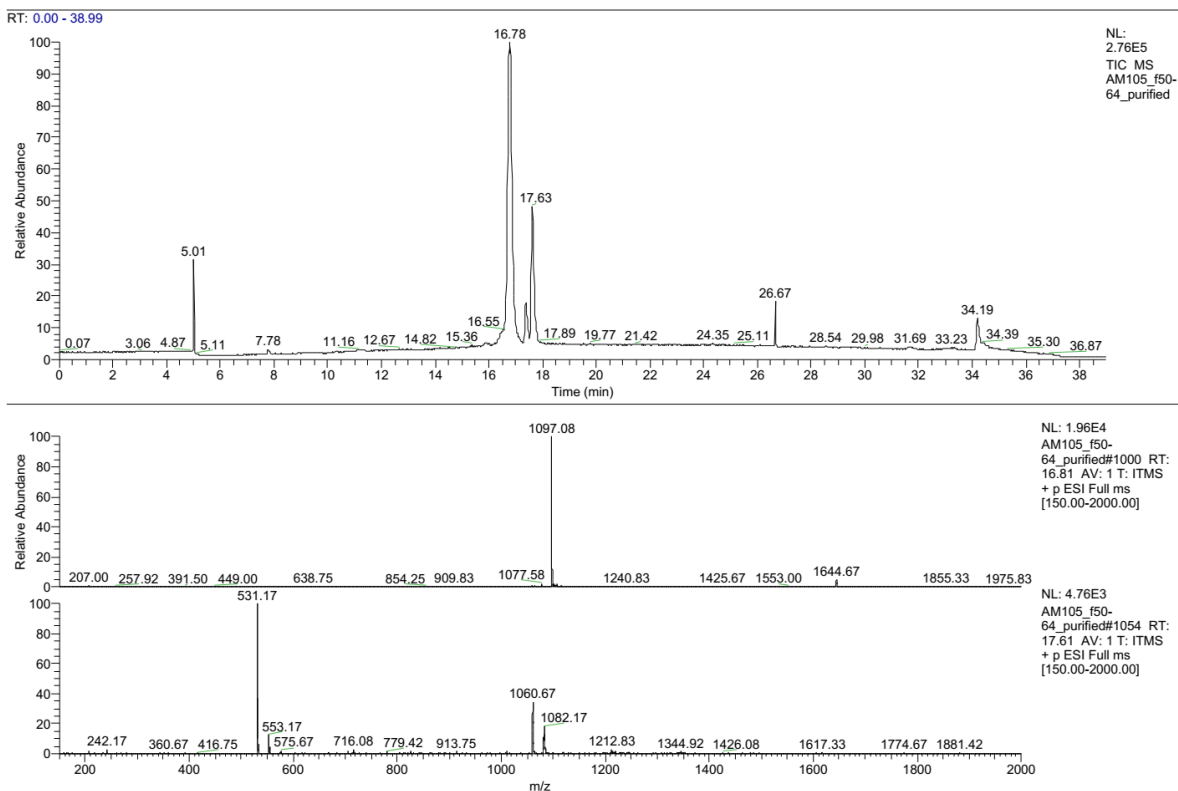


Figure 118 LCMS Cycloaddition of the second loop to obtain product 20.

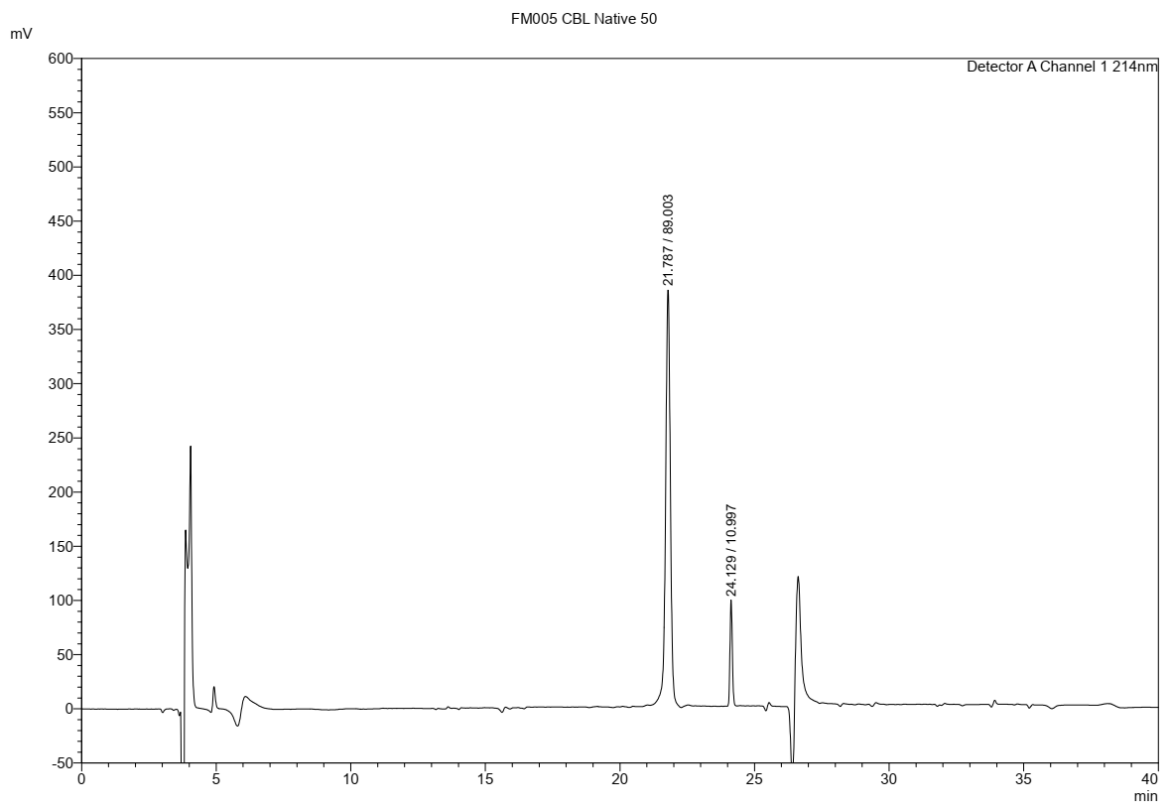


Figure 119 analytical HPLC of purified compound 20.

E:\LCMS\...AM109\_f20-15\_180430110455

30/04/2018 11:04:55

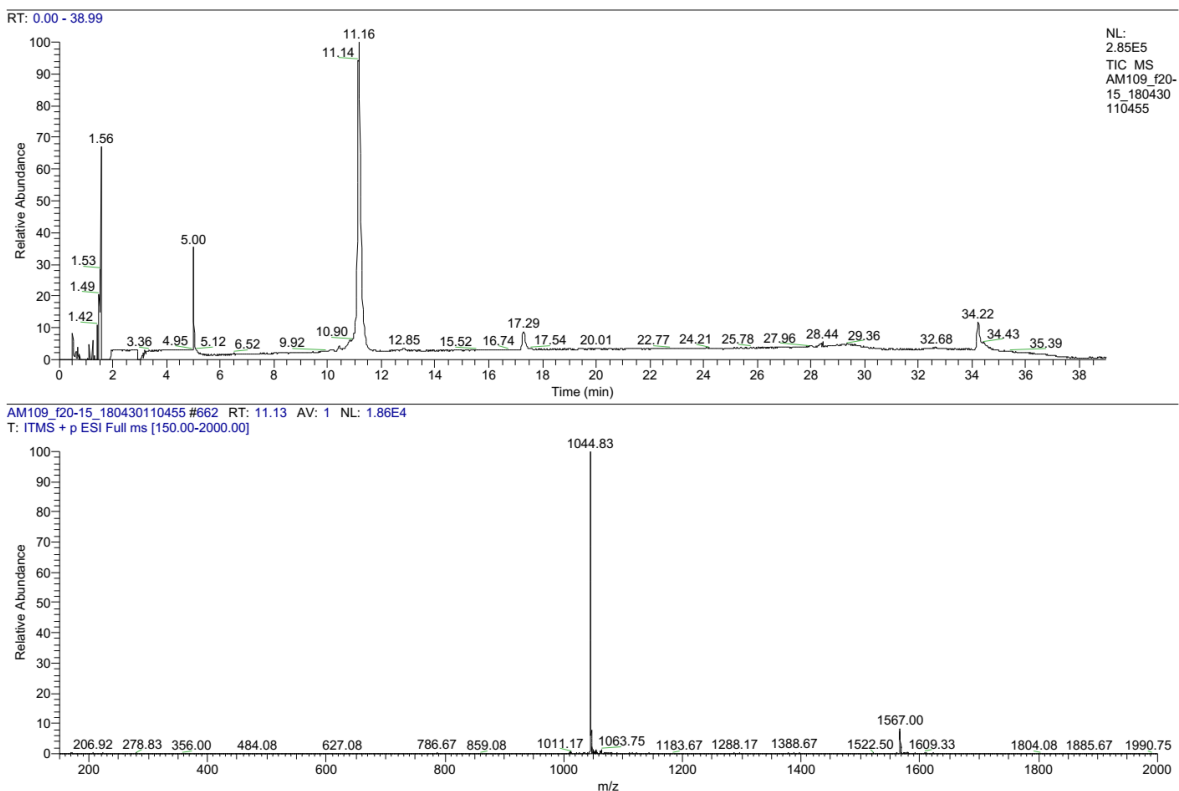


Figure 120 LCMS of product 21 after TIPS removal.

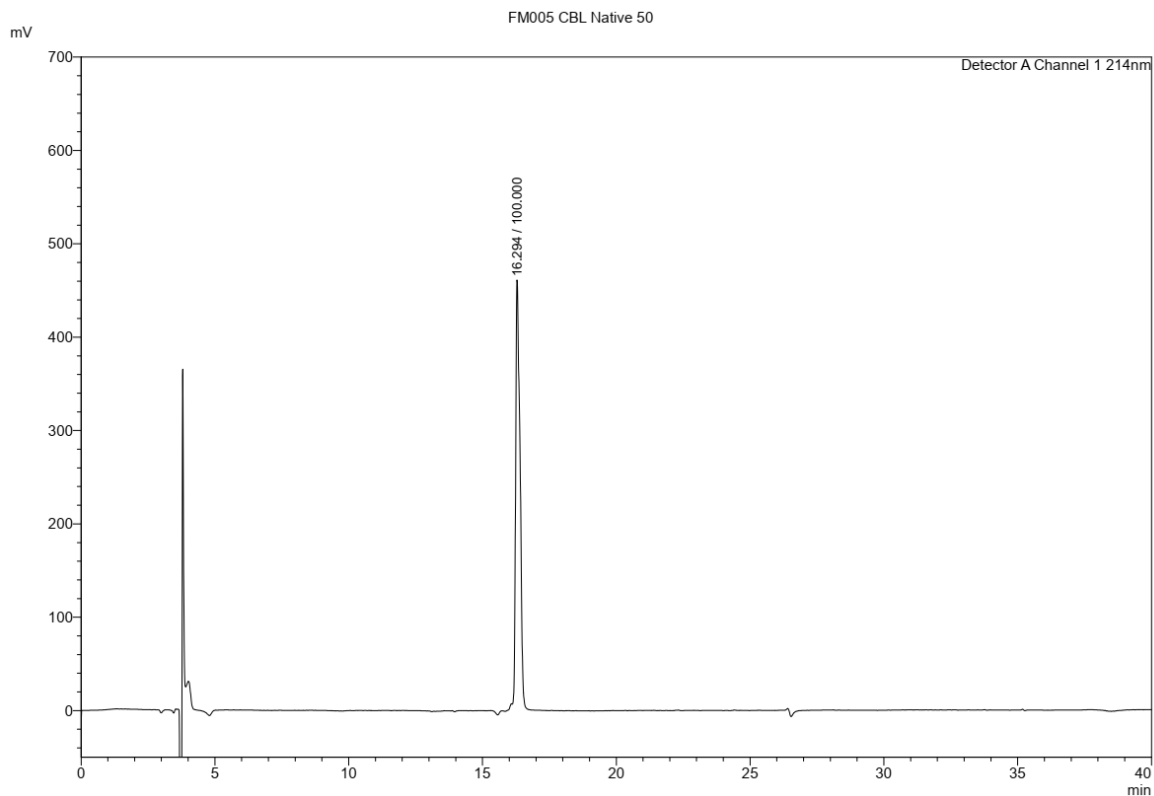


Figure 121 analytical HPLC of purified compound 21.

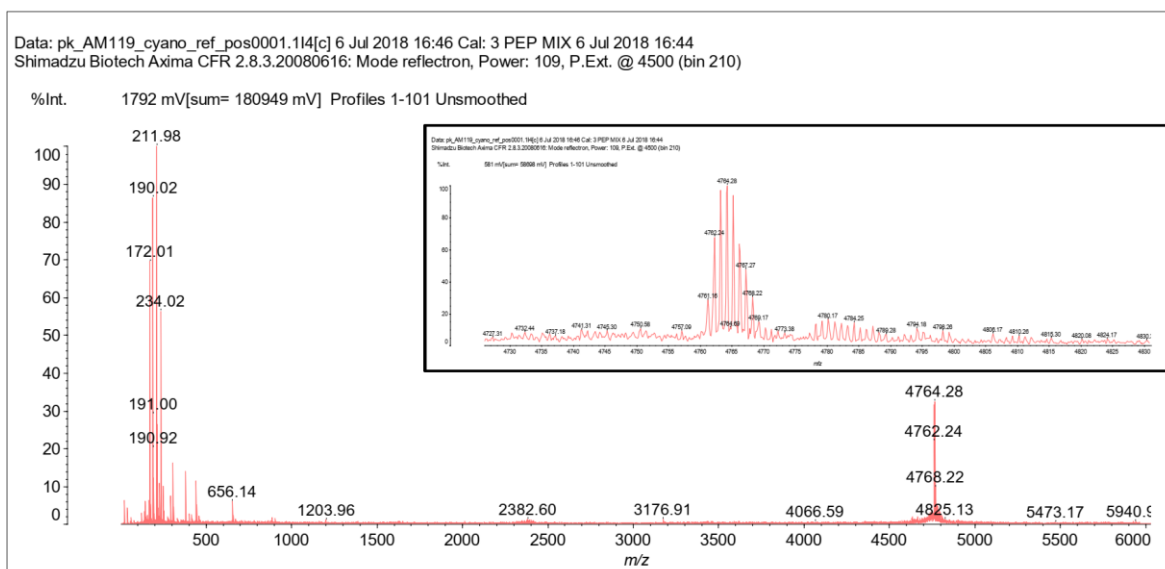


Figure 122 MALDI-TOF of compound 22.

E:\LCMS\...AM119\_f100

01/05/2018 09:09:01

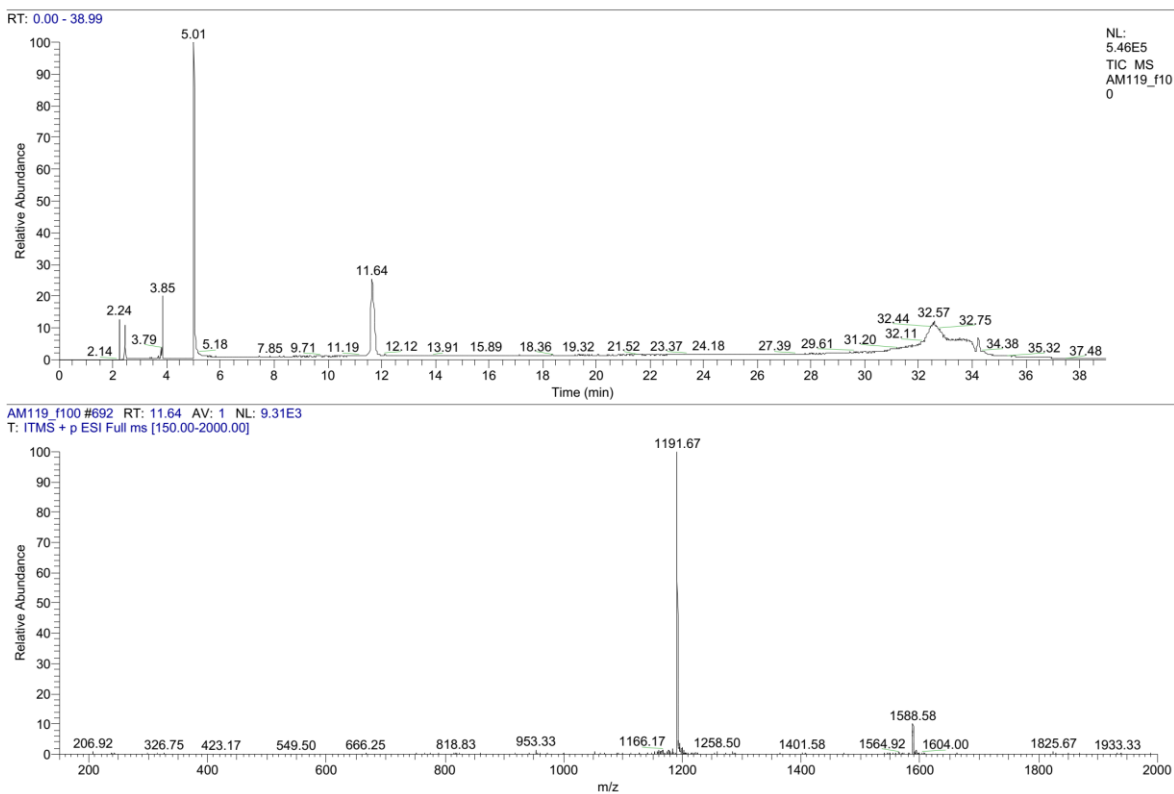


Figure 123 LCMS of compound 22.

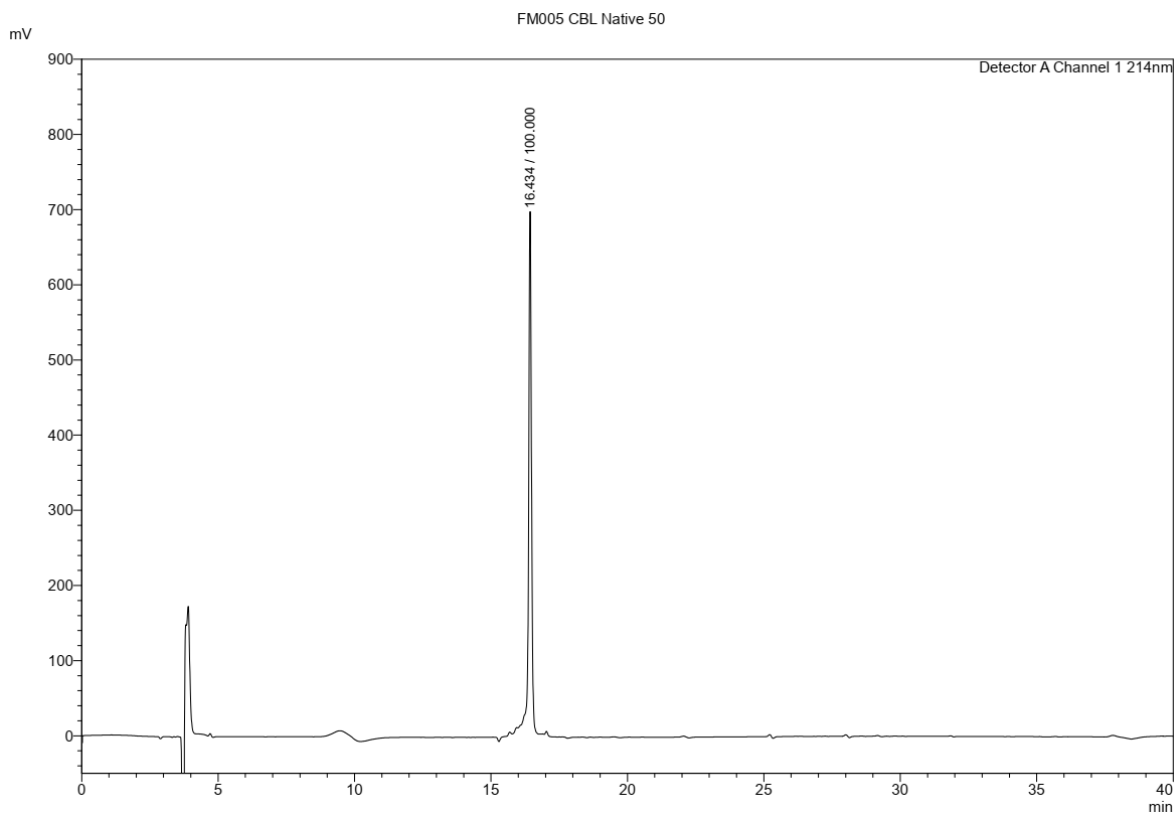


Figure 124 analytical HPLC of purified compound 22.

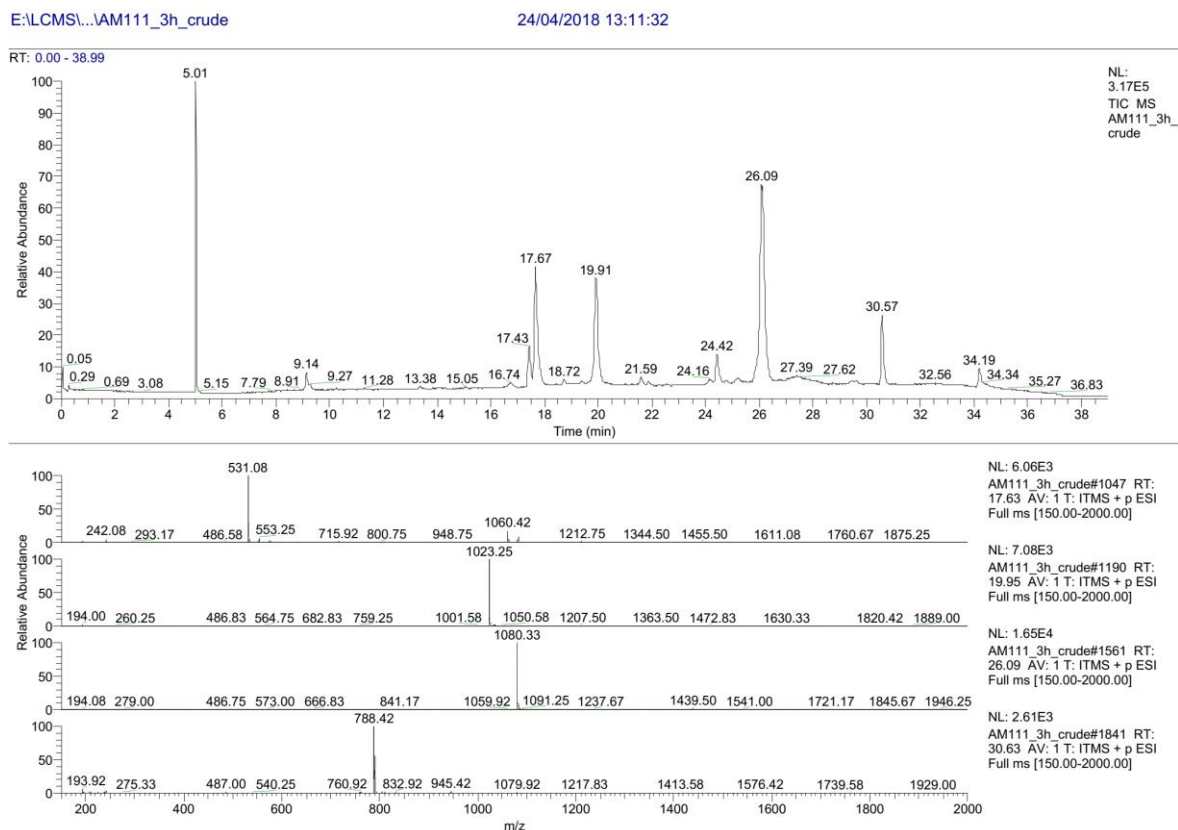


Figure 125 LCMS of cycloaddition of the first loop in synthesis of compound 30.

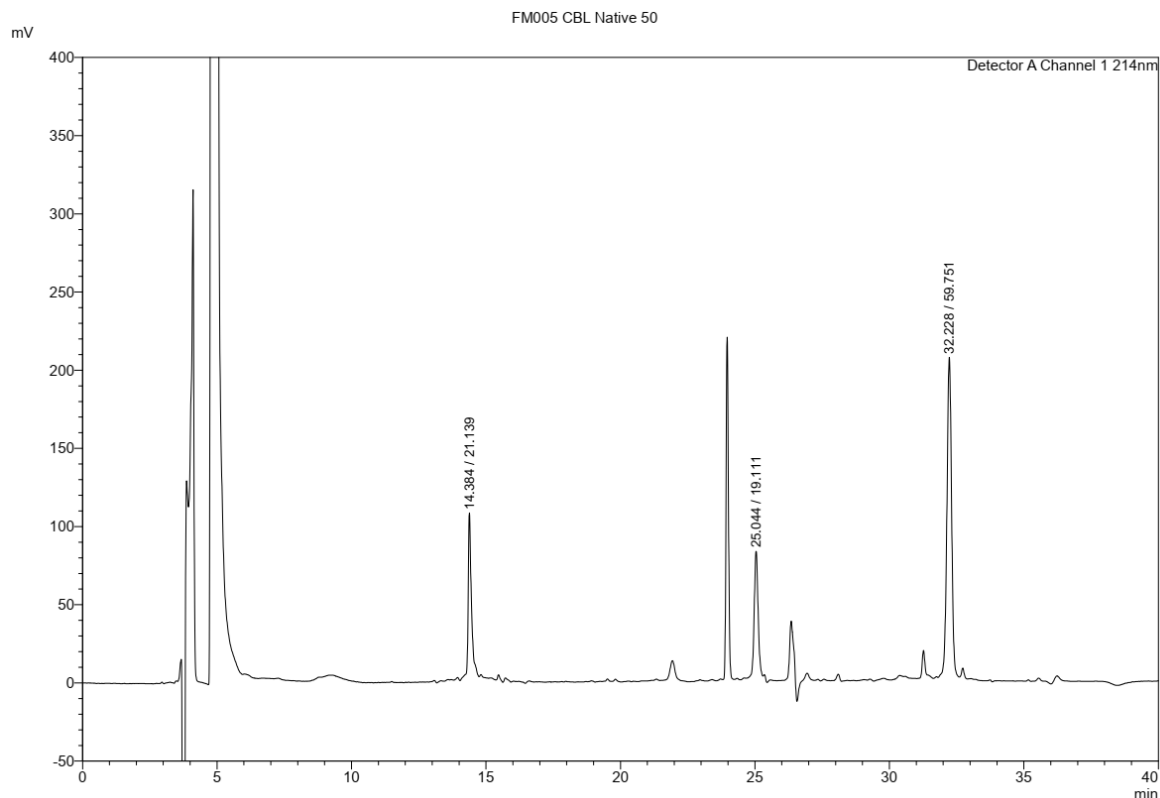


Figure 126 analytical HPLC of crude reaction mixture after cycloaddition of the first loop in the synthesis of compound 30.

E:\LCMS\...AM112\_1h\_crude\_180424152937

24/04/2018 15:29:37

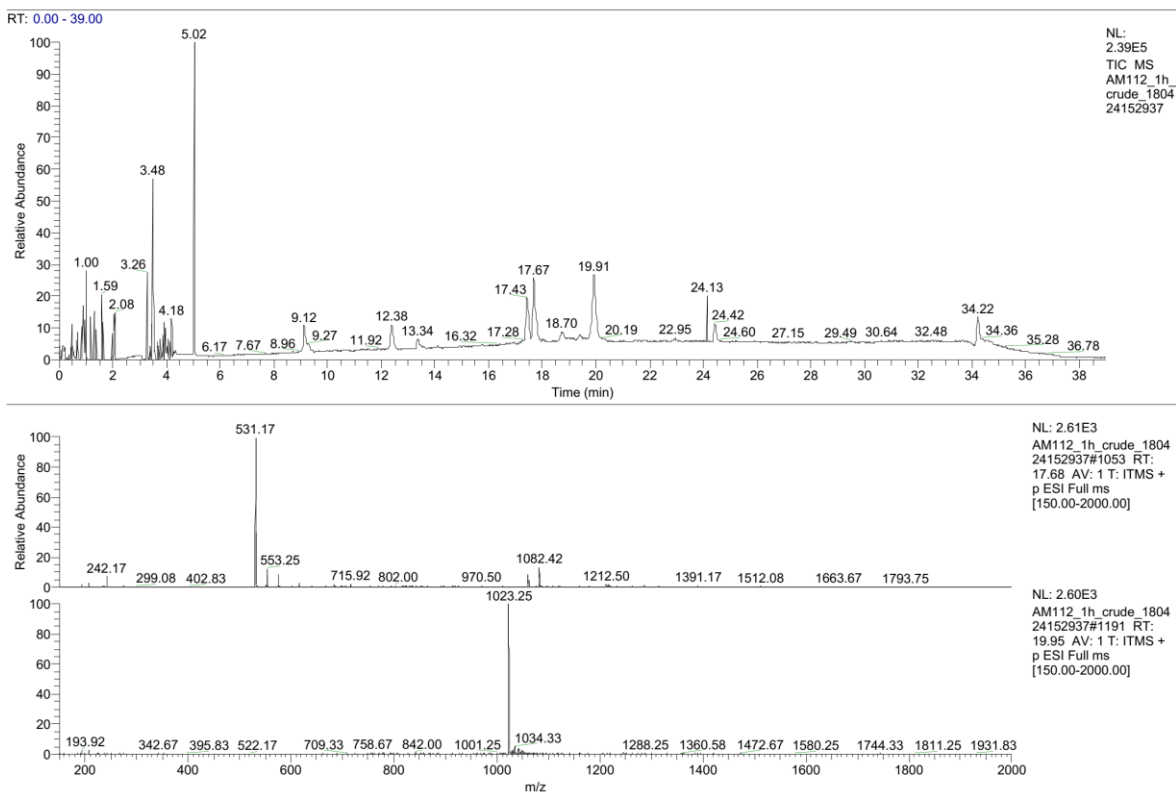


Figure 127 LCMS of TES removal to obtain compound 27.

E:\LCMS\...am113\_F91-87

30/04/2018 10:22:35

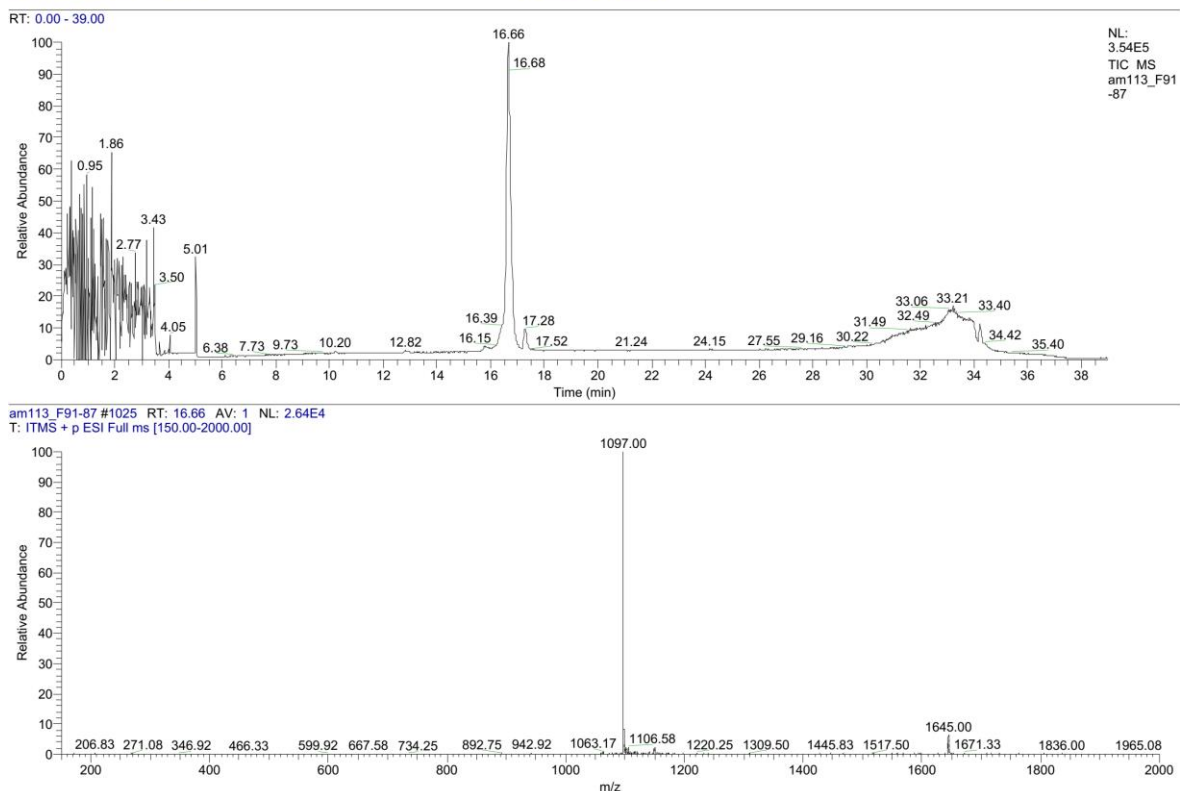


Figure 128 LCMS of product 28 after cycloaddition of the second loop.

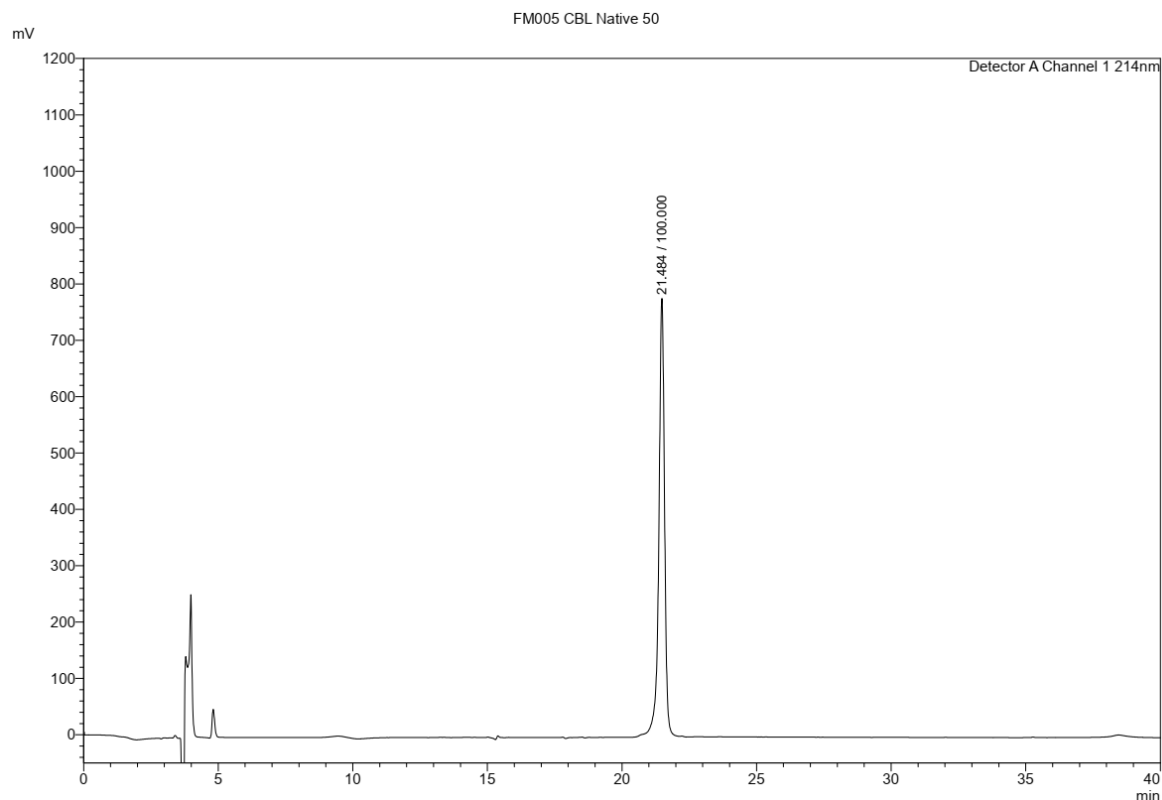


Figure 129 analytical HPLC of purified product 28.

E:\LCMS\...AM122\_f70-61

08/05/2018 12:34:06

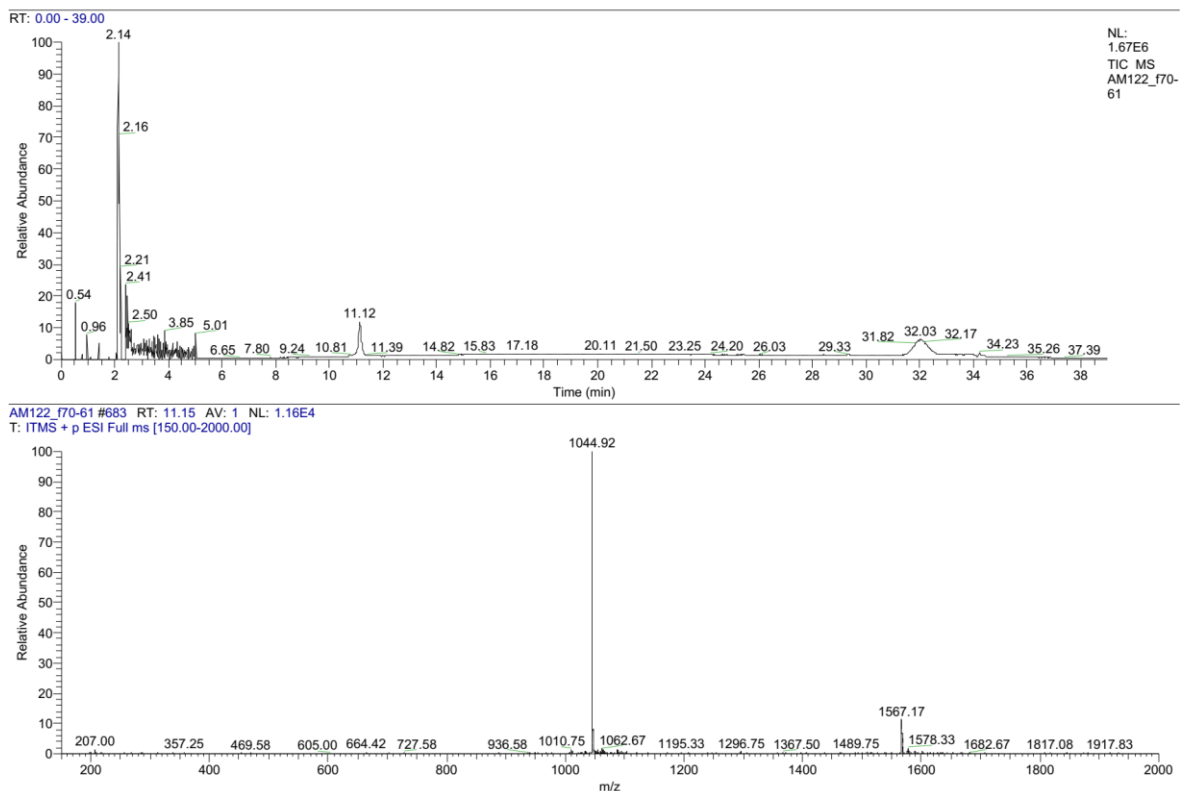


Figure 130 LCMS of product 29 after TIPS removal.

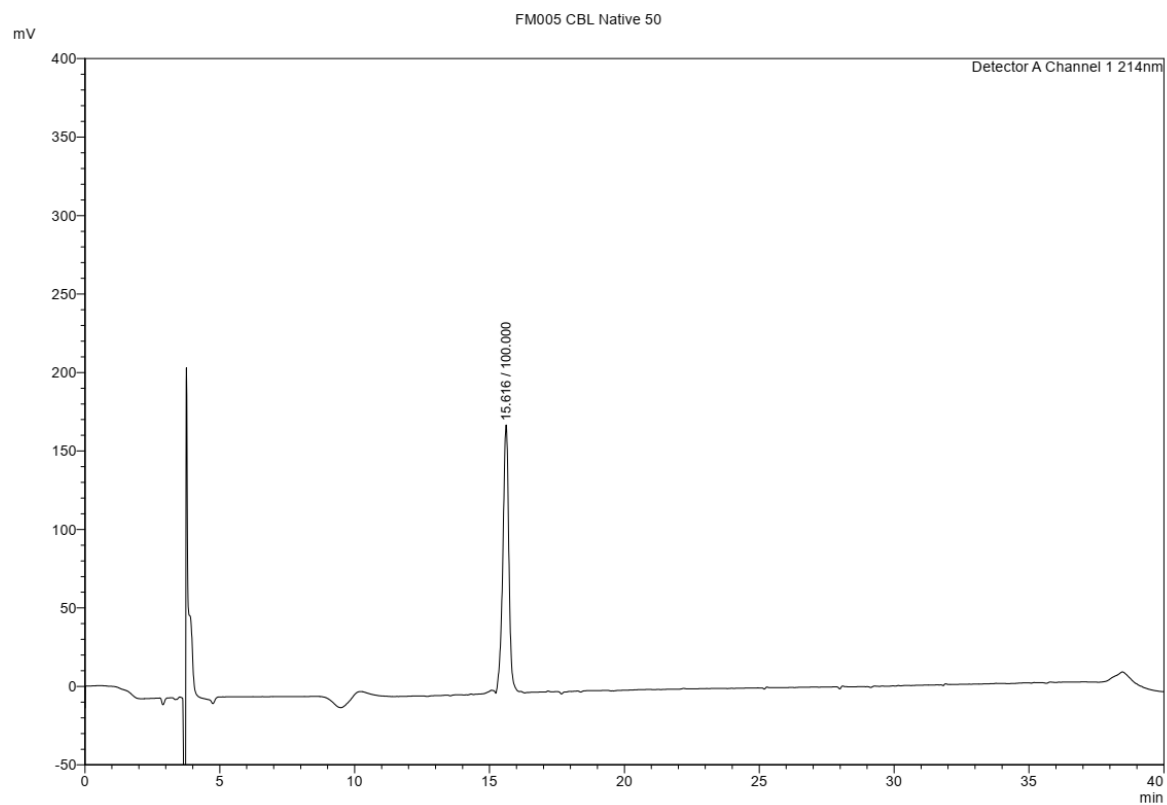


Figure 131 analytical HPLC product 29 after TIPS removal.



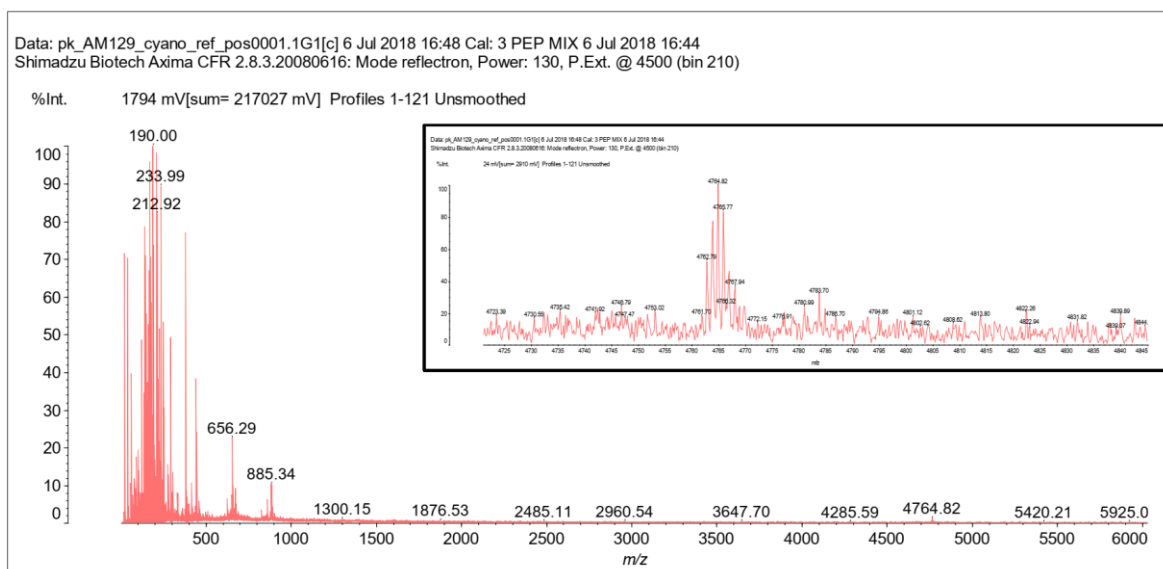


Figure 132 MALDI-TOF of product 30.

E:\LCMS\...\AM129\_f88-85

12/06/2018 08:56:23

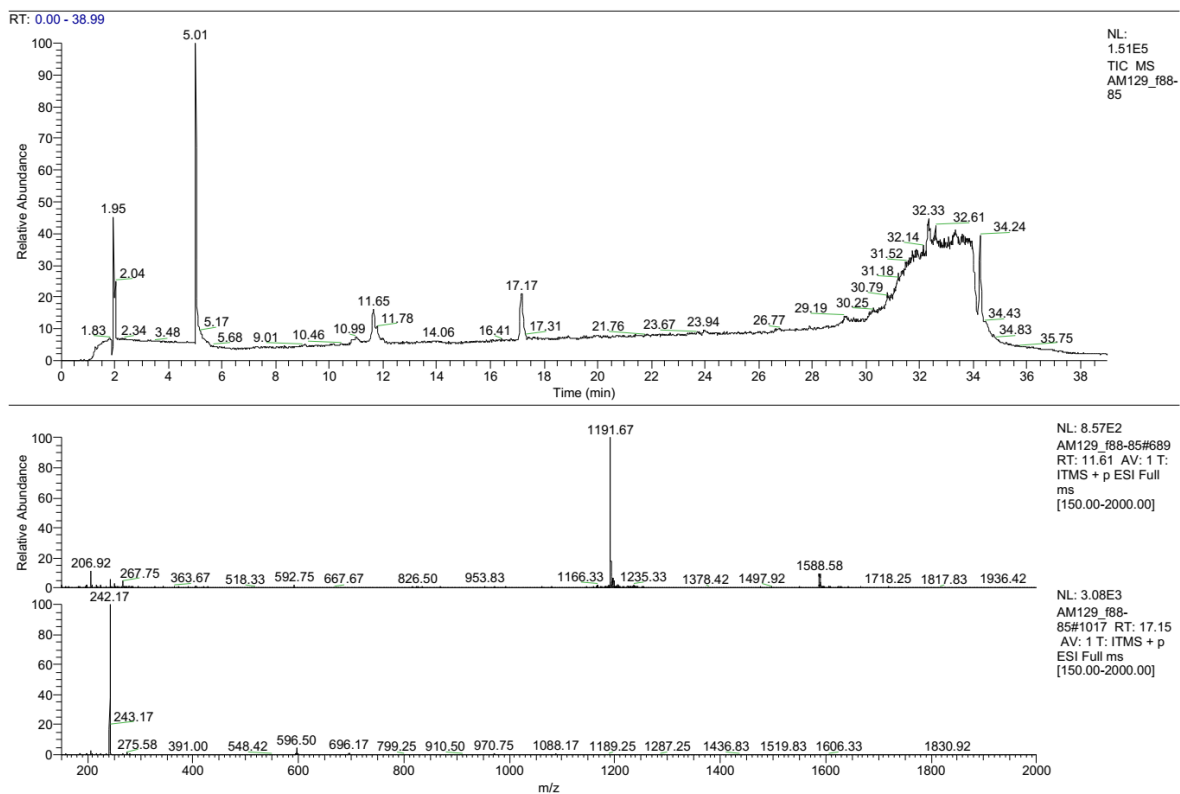


Figure 133 LCMS of product 30.

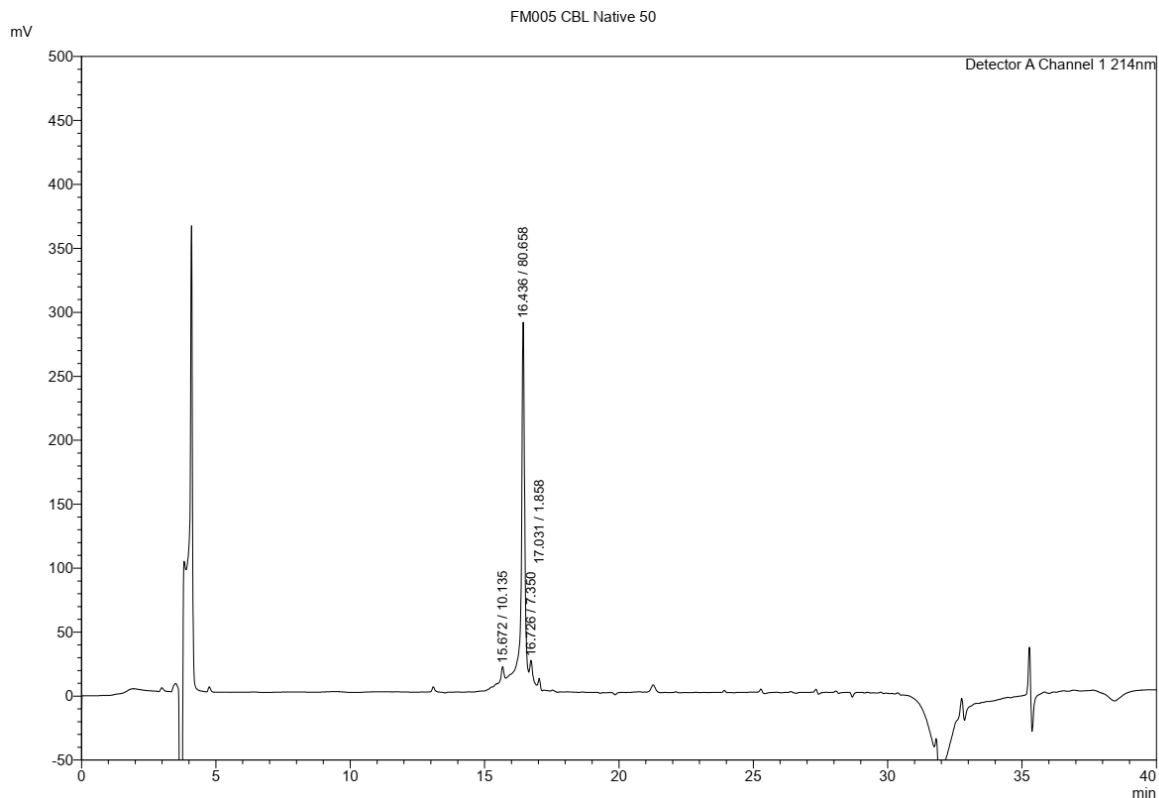


Figure 134 analytical HPLC of purified compound 30.

E:\LCMS\...AM156\_2hcrude\_40min

06/06/2018 13:23:32

RT: 0.00 - 38.99

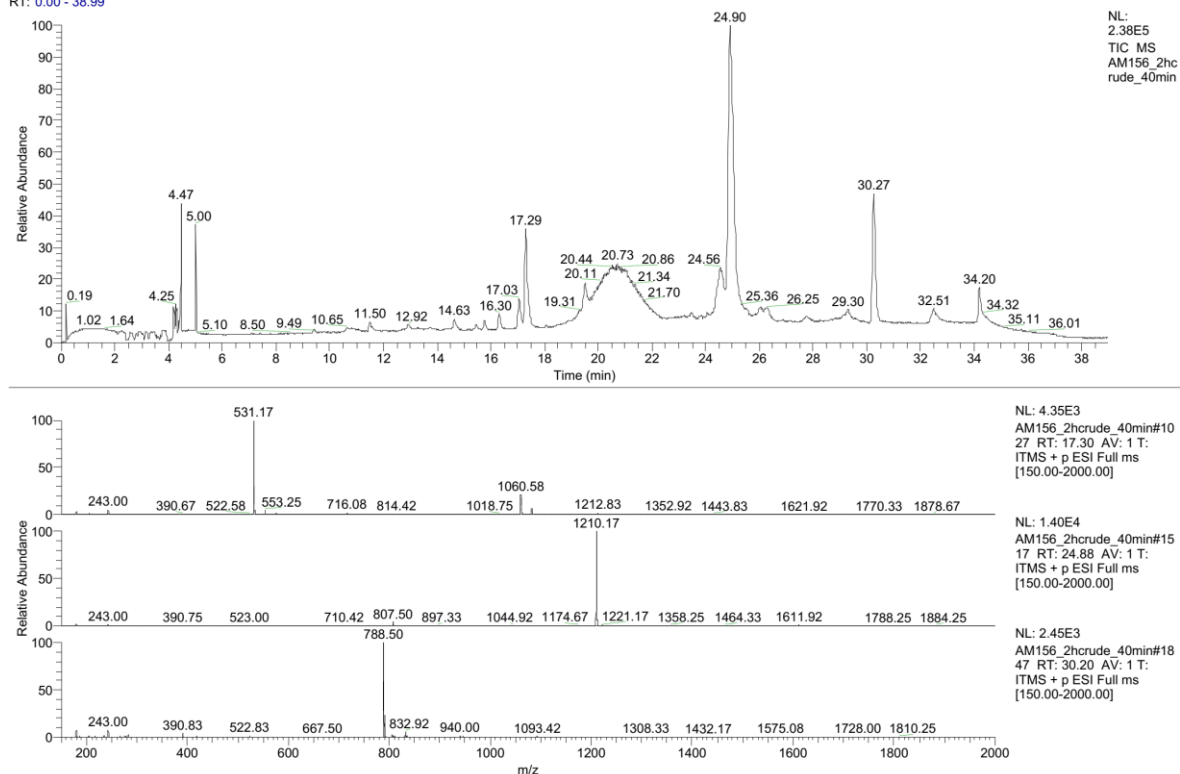


Figure 135 LCMS after cycloaddition of first loop in the synthesis of compound 26.

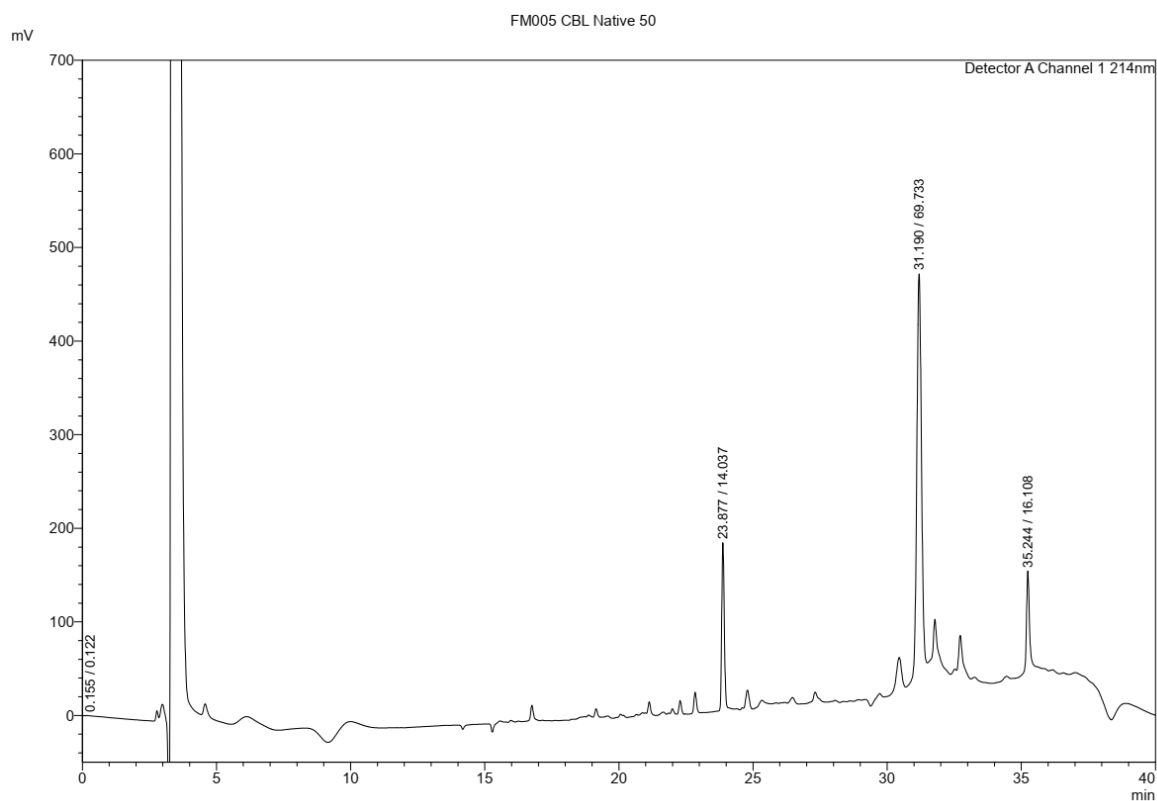


Figure 136 analytical HPLC after cycloaddition of the first loop in the synthesis of compound 26.

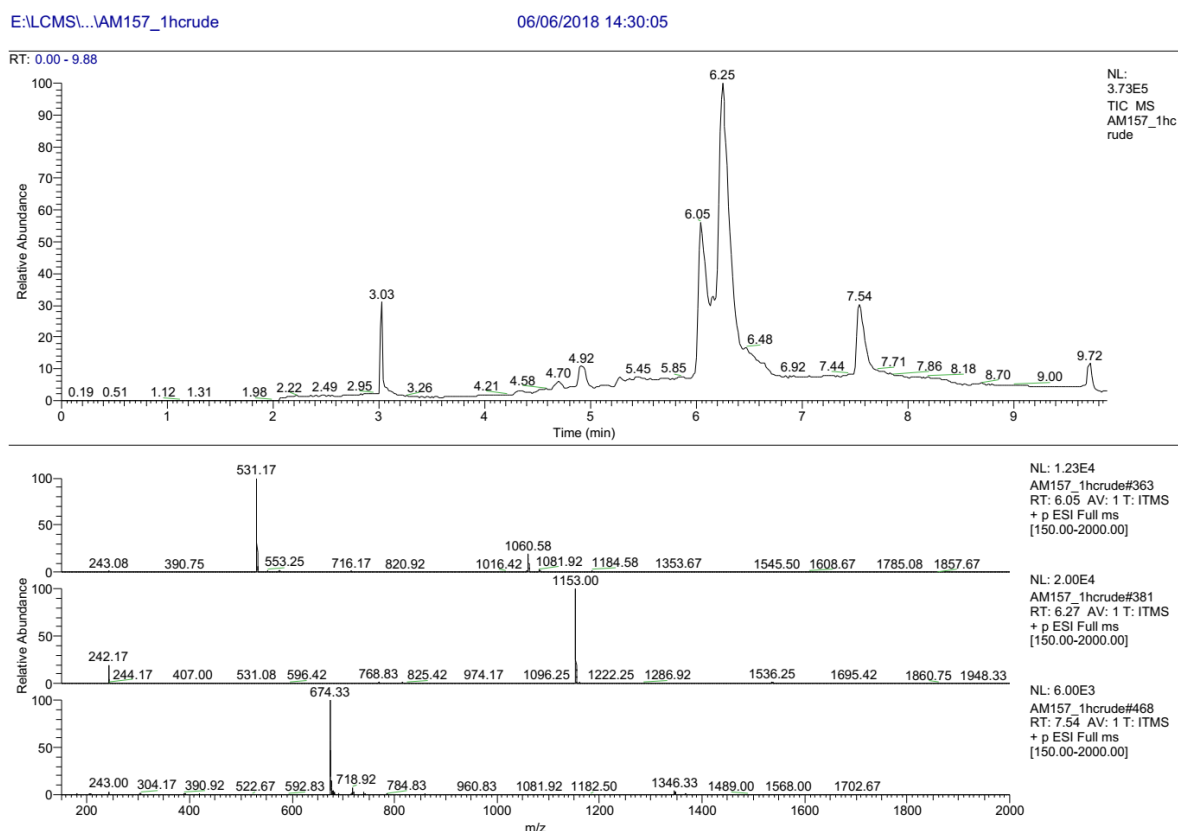
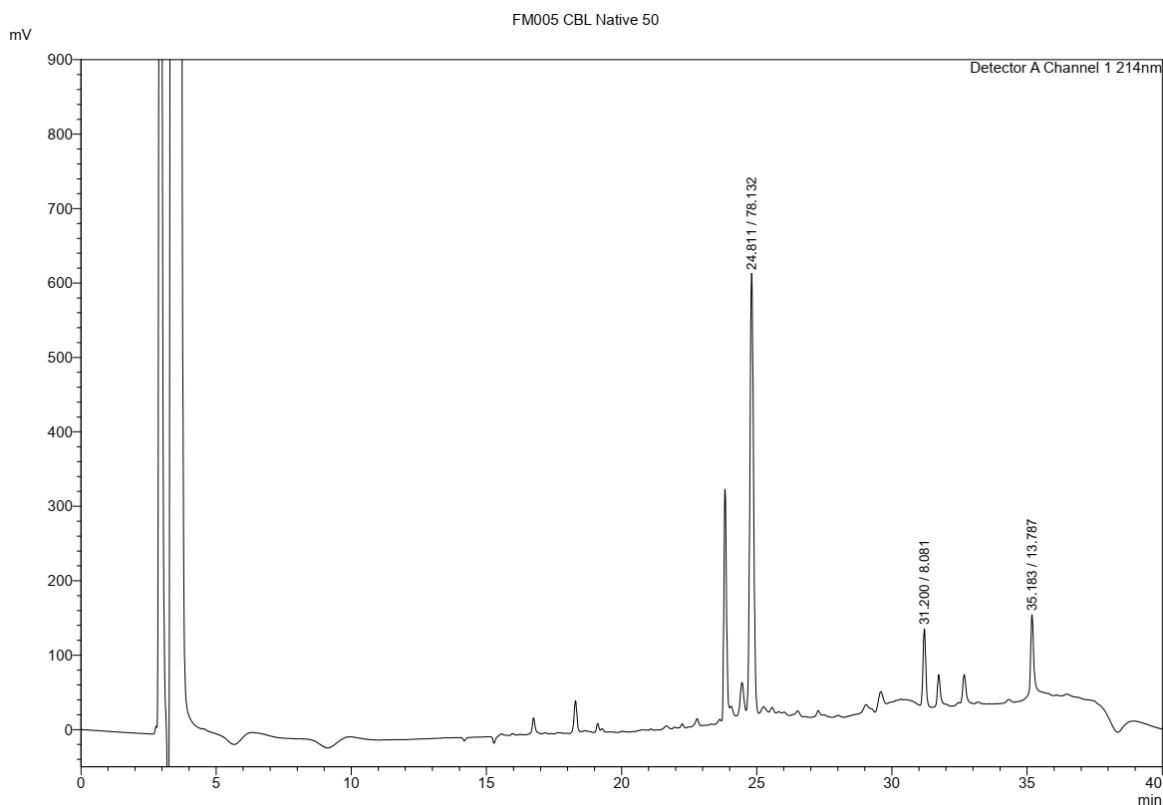


Figure 137 LCMS after TES removal to obtain compound 23.



E:\LCMS\...AM158\_f22-15\_40min

07/06/2018 09:12:26

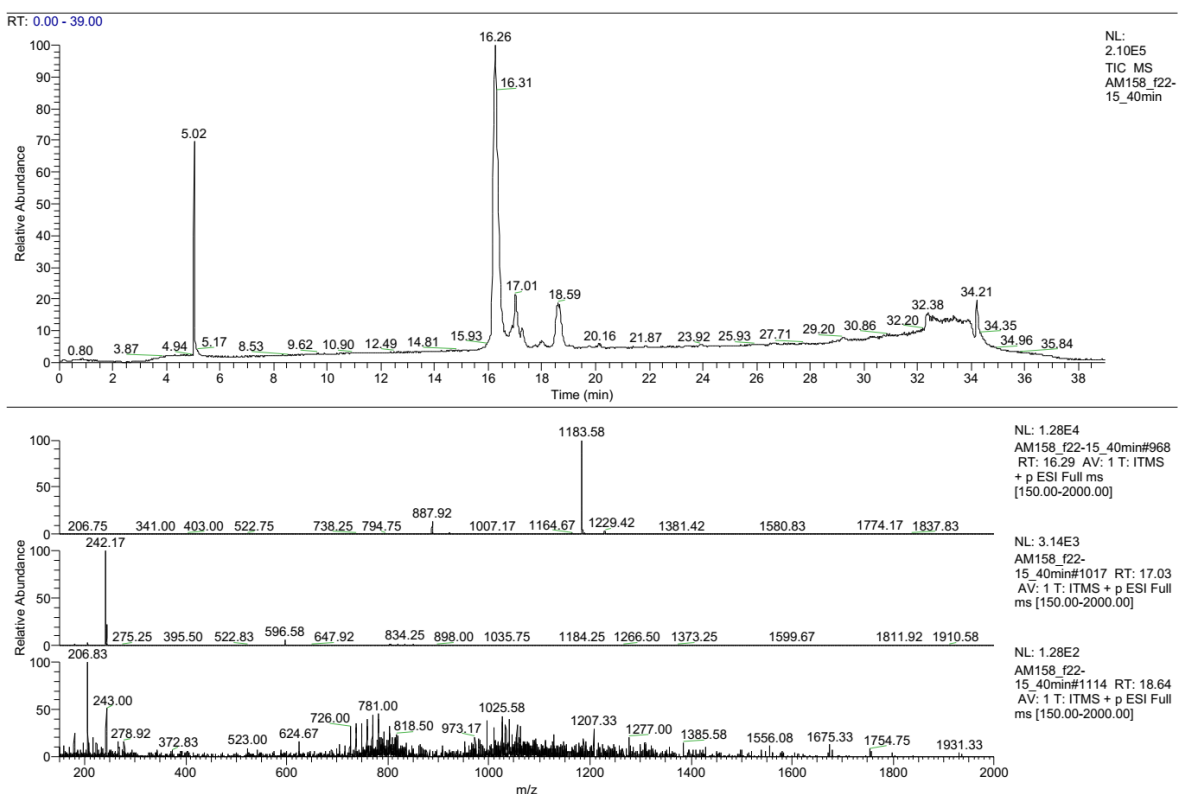


Figure 139 LCMS after cycloaddition of the second loop to obtain product 24.

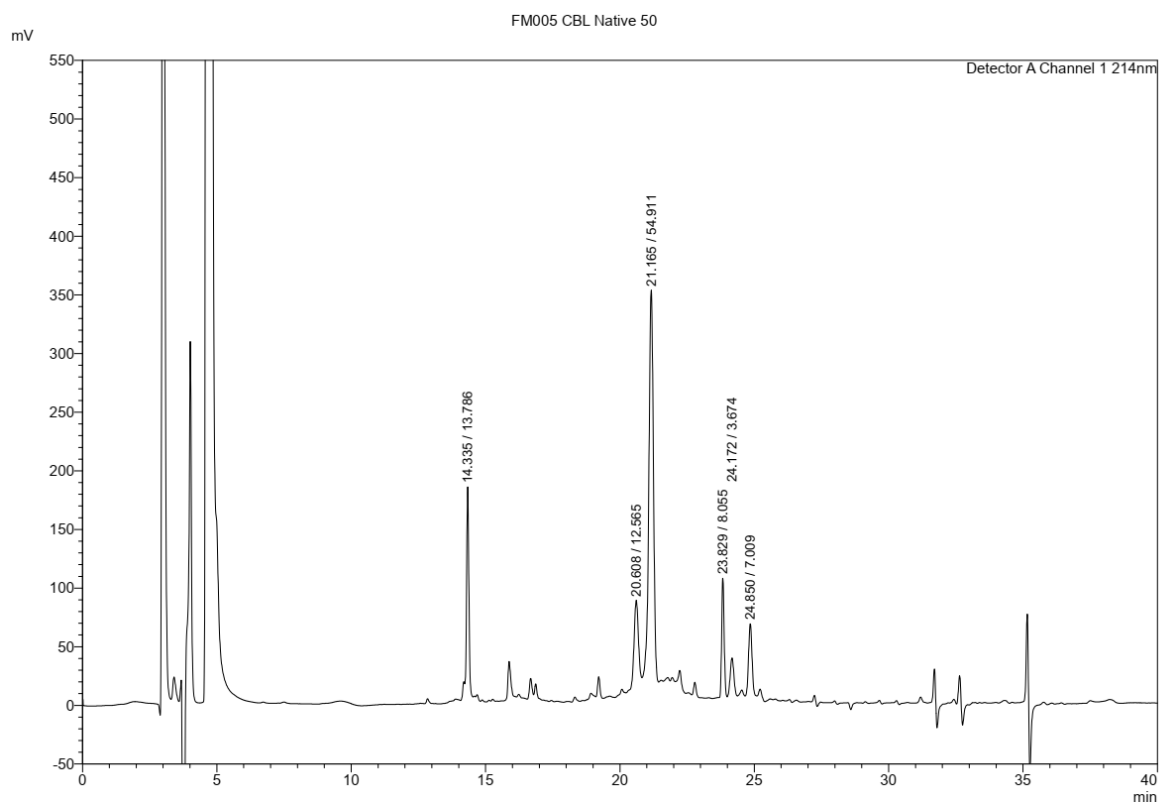


Figure 140 analytical HPLC of crude reaction mixture to obtain product 24.

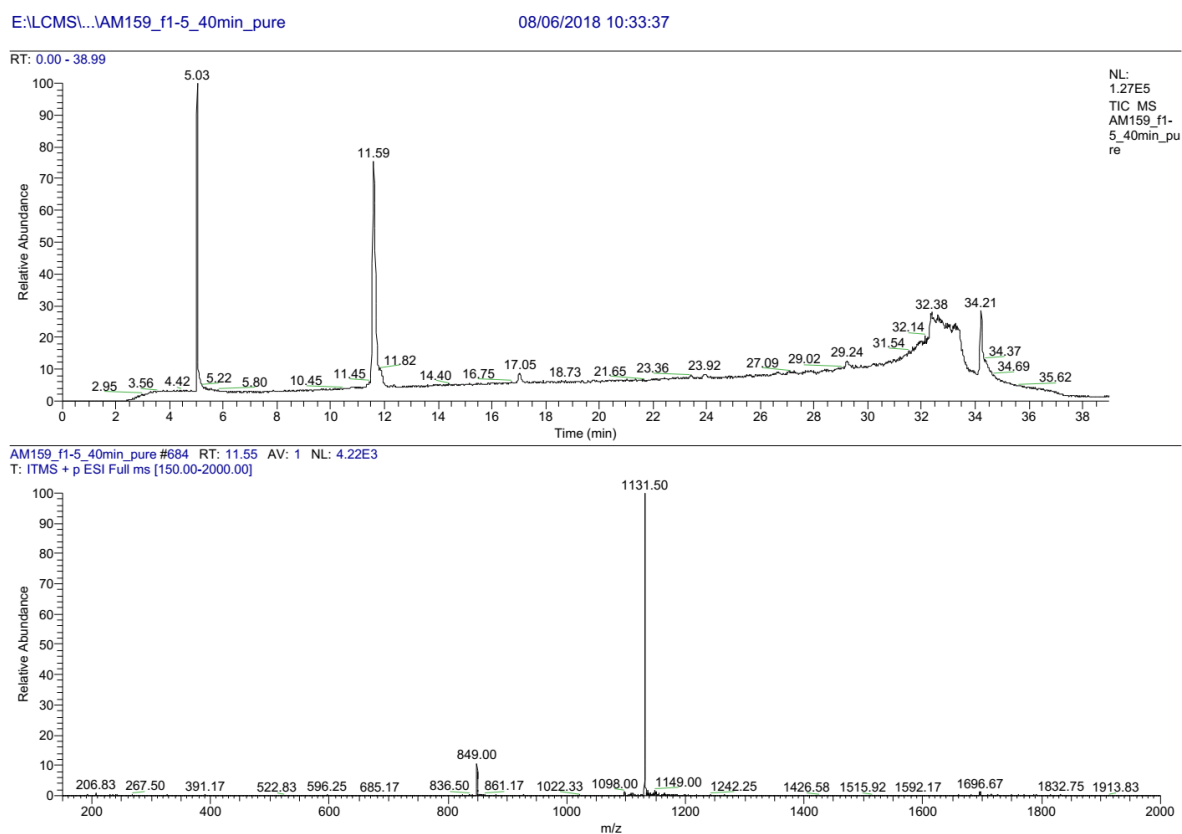


Figure 141 LCMS of product 25 after TIPS removal.

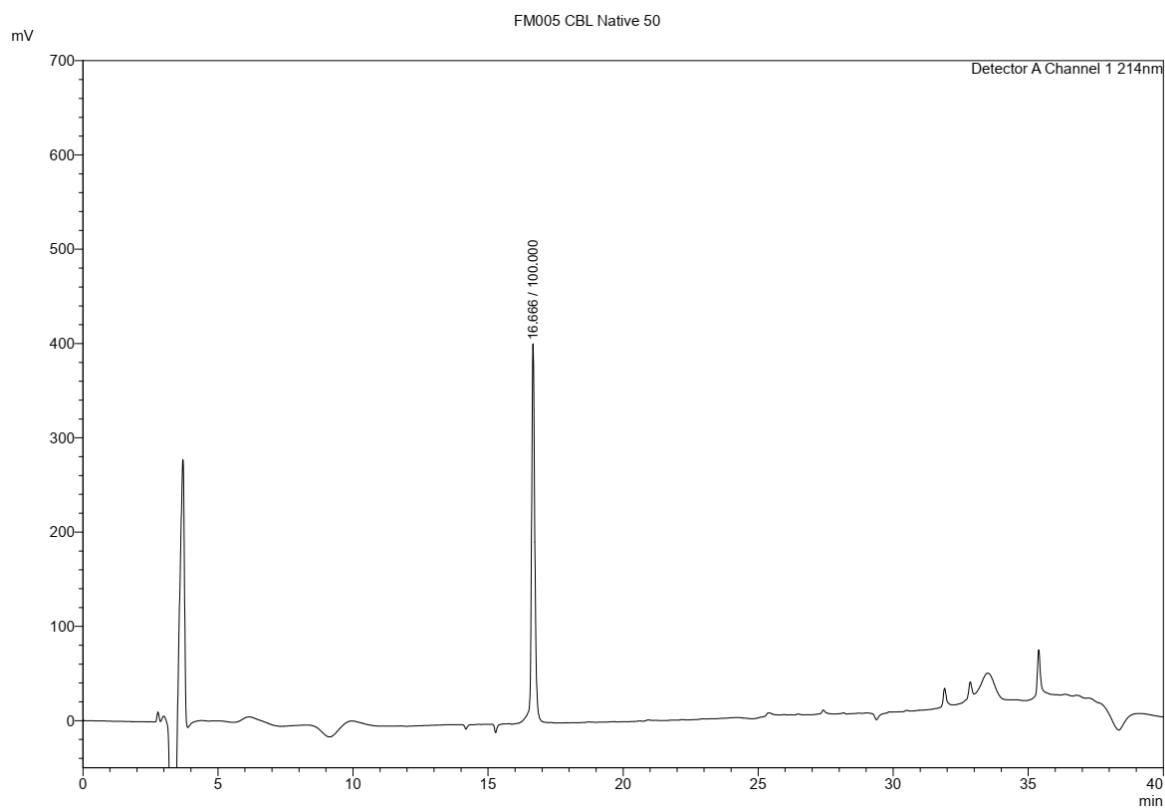


Figure 142 analytical HPLC of purified compound 25 after TIPS removal.

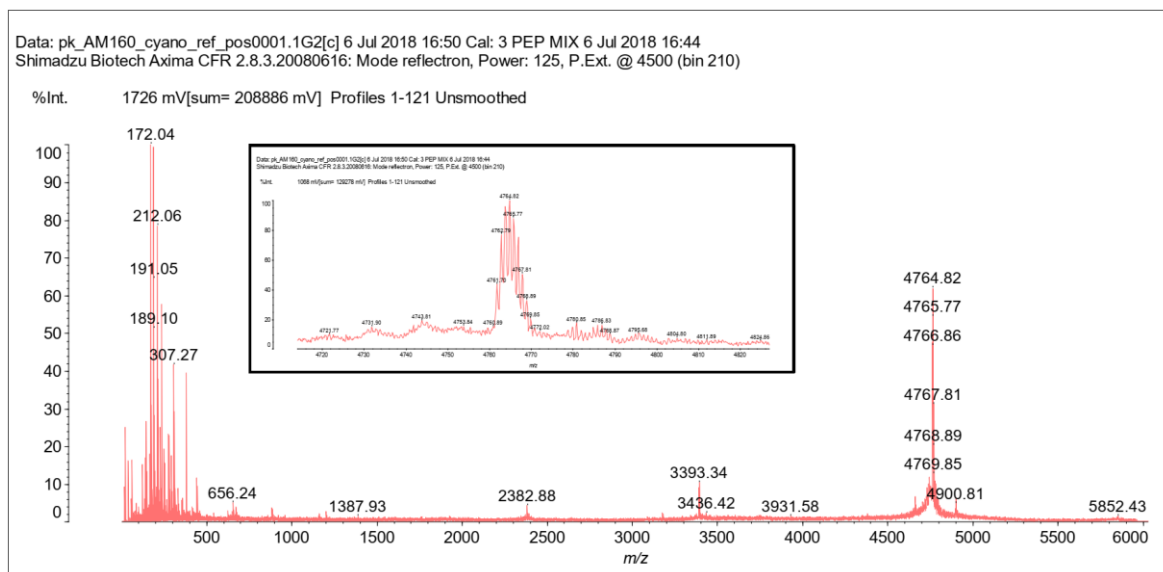


Figure 143 MALDI-TOF of product 26.

E:\LCMS\...AM160\_f6-10

12/06/2018 09:38:49

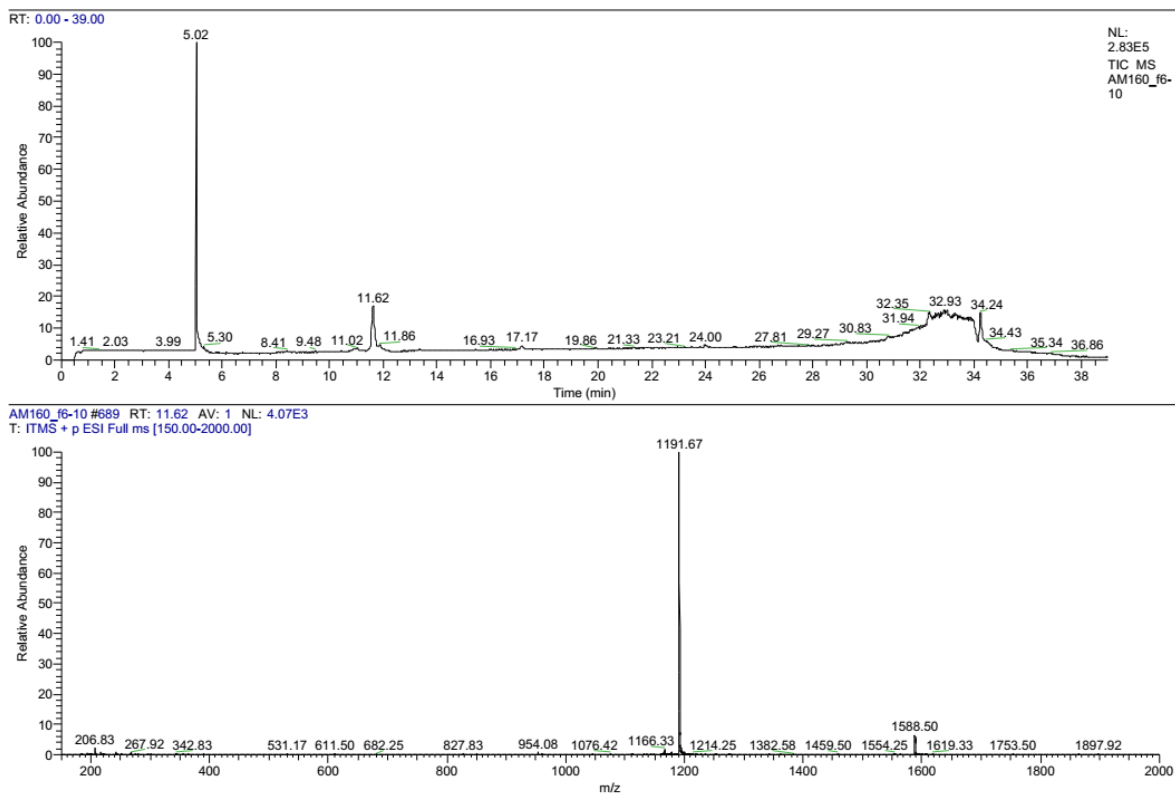


Figure 144 LCMS of product 26.

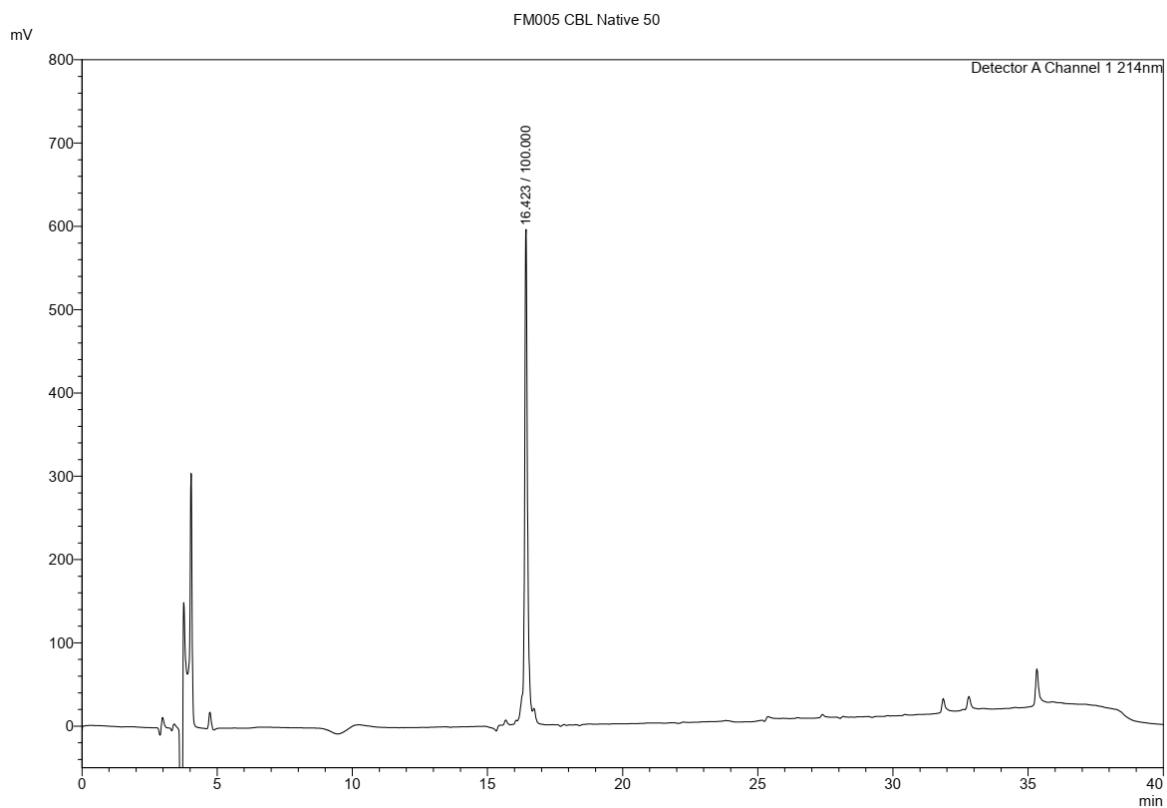


Figure 145 analytical HPLC of purified compound 26.

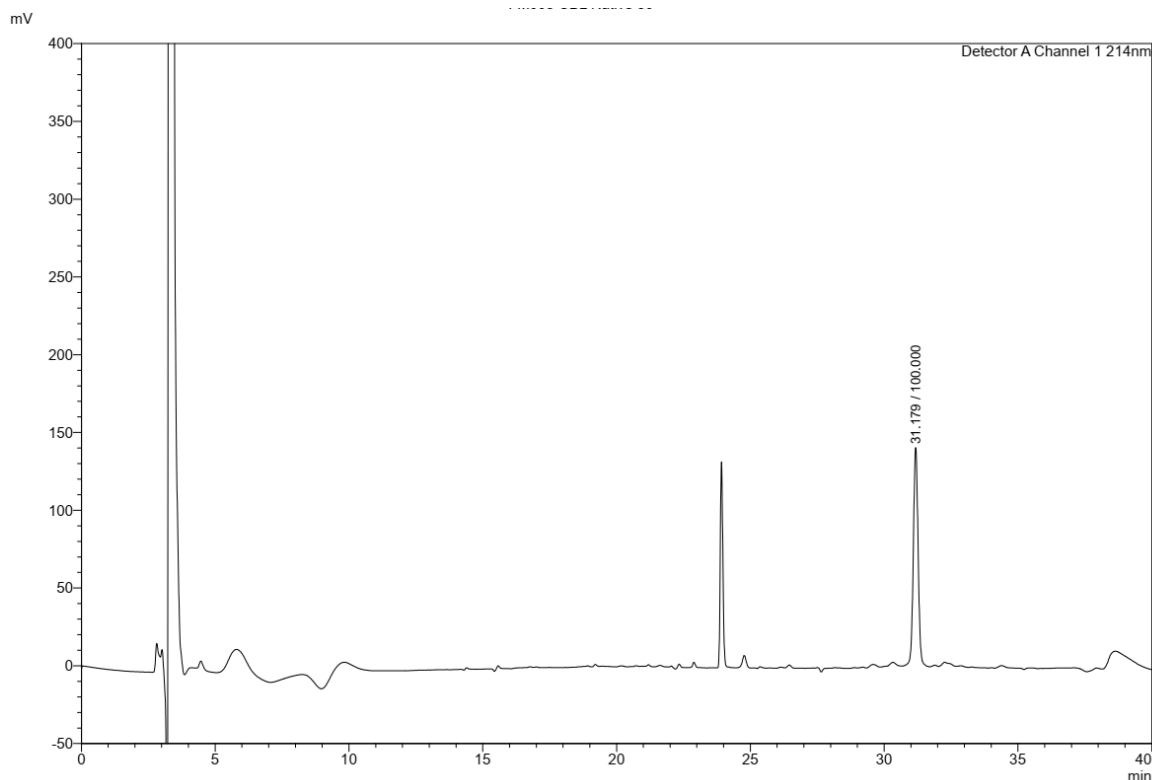


Figure 146 analytical HPLC: attachment of loop 2 to the TAC scaffold in the synthesis of compound 26 with reverse order of loops attachment.

E:\LCMS\...AM136\_2hcrude\_click1\_40min

20/05/2018 14:14:59

RT: 0.00 - 38.99

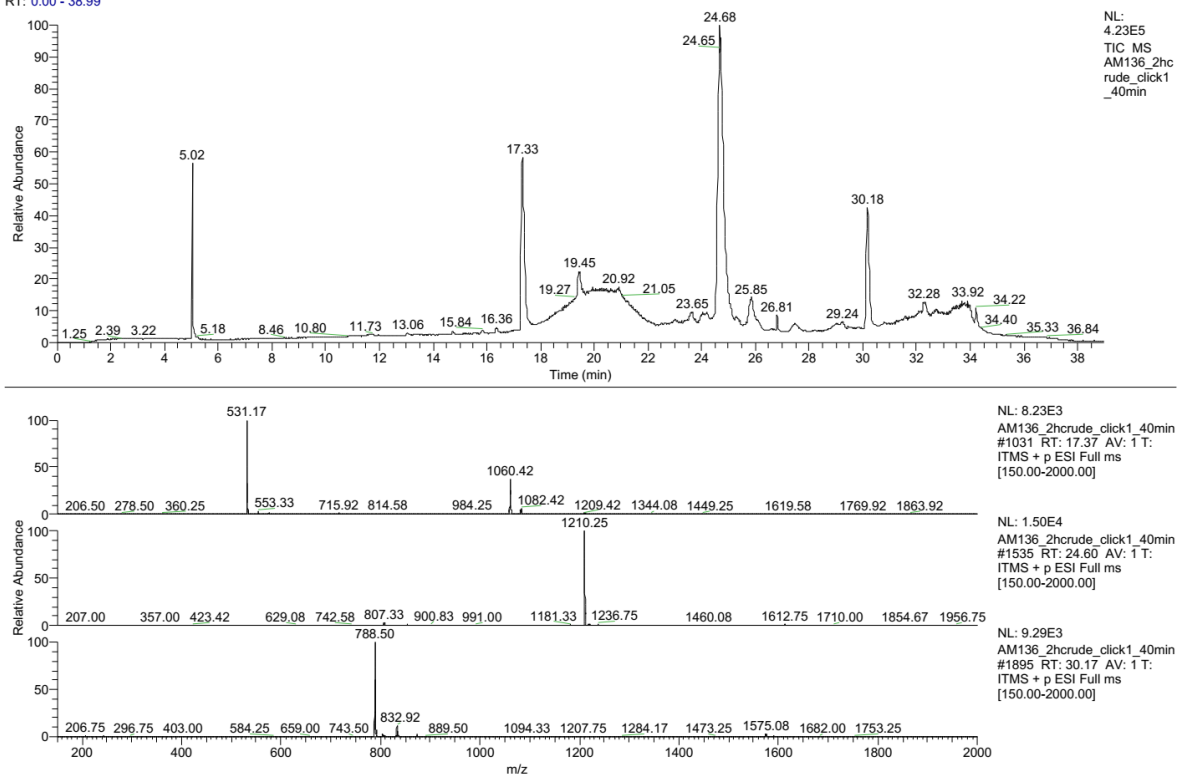


Figure 147 LCMS analysis of attachment of loop 2 to the TAC scaffold in the synthesis of compound 26 with reversed order of loops attachment.



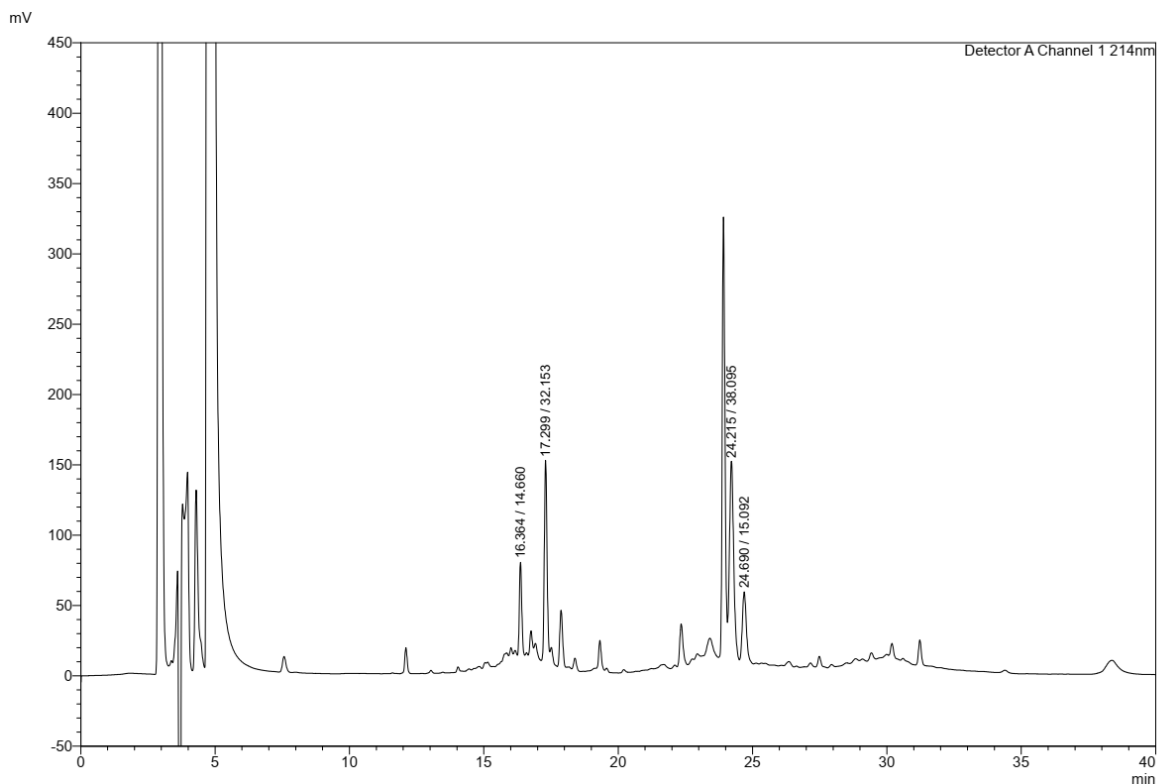


Figure 148 TES deprotection to obtain product 22 in the synthesis of compound 26 with reverse order of loops attachment.

E:\LCMS\...AM137\_1hcrude\_40min

20/05/2018 15:55:20

RT: 0.00 - 39.00

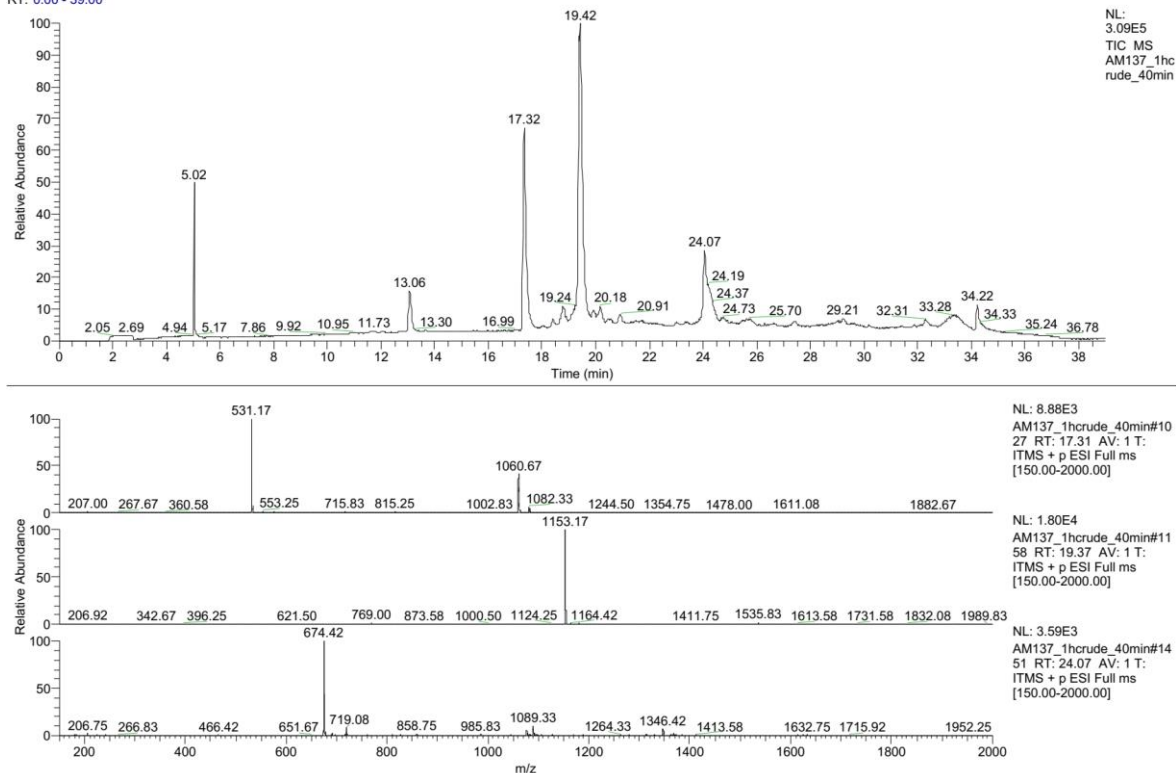


Figure 149 LCMS analysis of TES deprotection to obtain product 22 in the synthesis of compound 26 with reverse order of loops attachment.

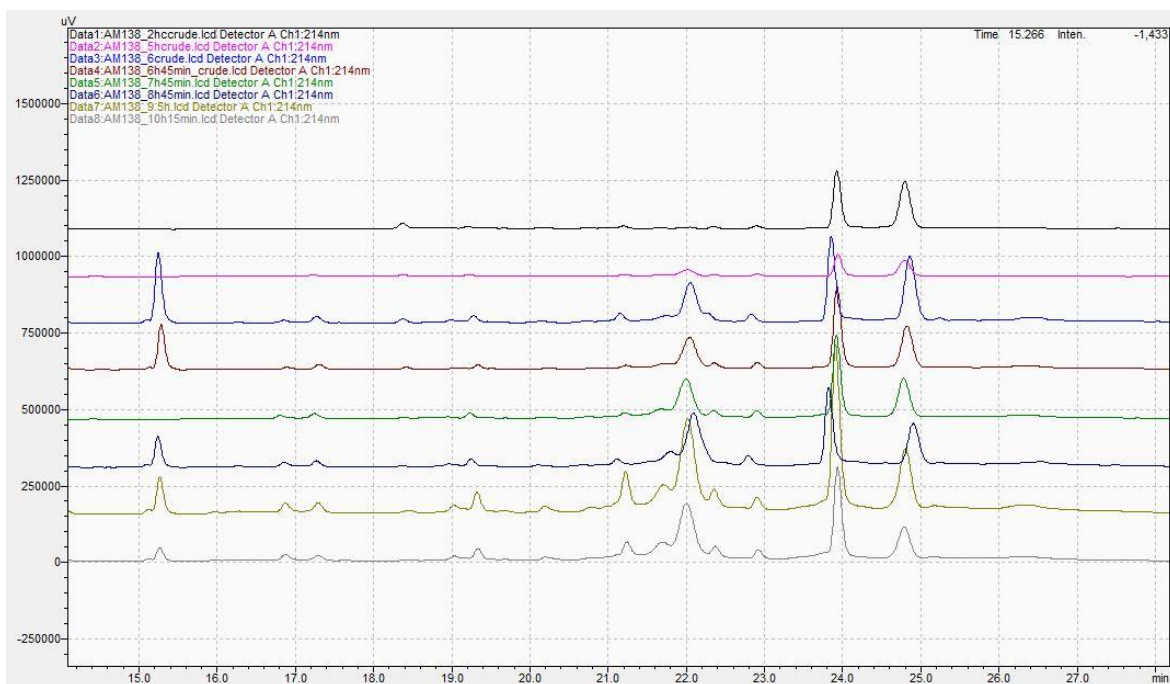


Figure 150 Progress of attachment of loop 3 (18) to compound 22 to obtain compound 31 in the synthesis of compound 26 with reverse order of loops attachment as analysed by analytical HPLC.

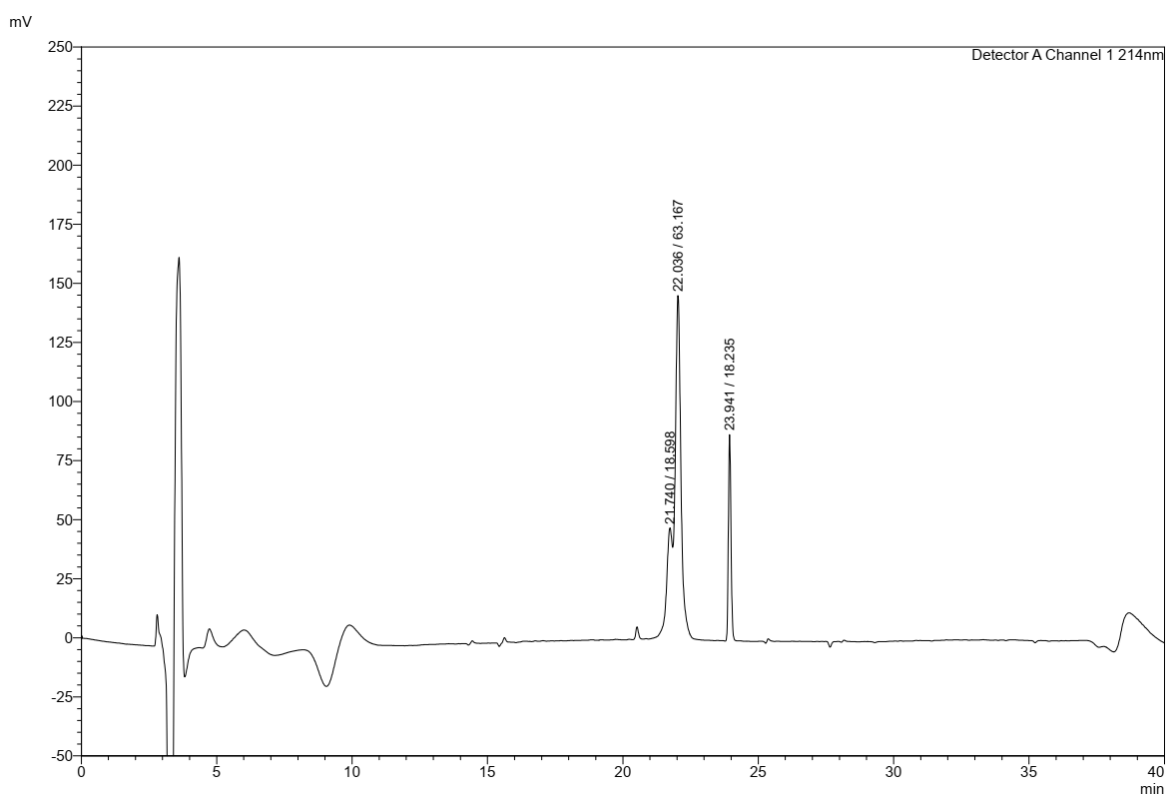


Figure 151 analytical HPLC of purified compound 31 (in the synthesis of compound 26 with reverse order of loops attachment as analysed by analytical HPLC), which could not be separated from TBTA.

E:\LCMS\...AM138\_aftprep\_40min

22/05/2018 09:29:17

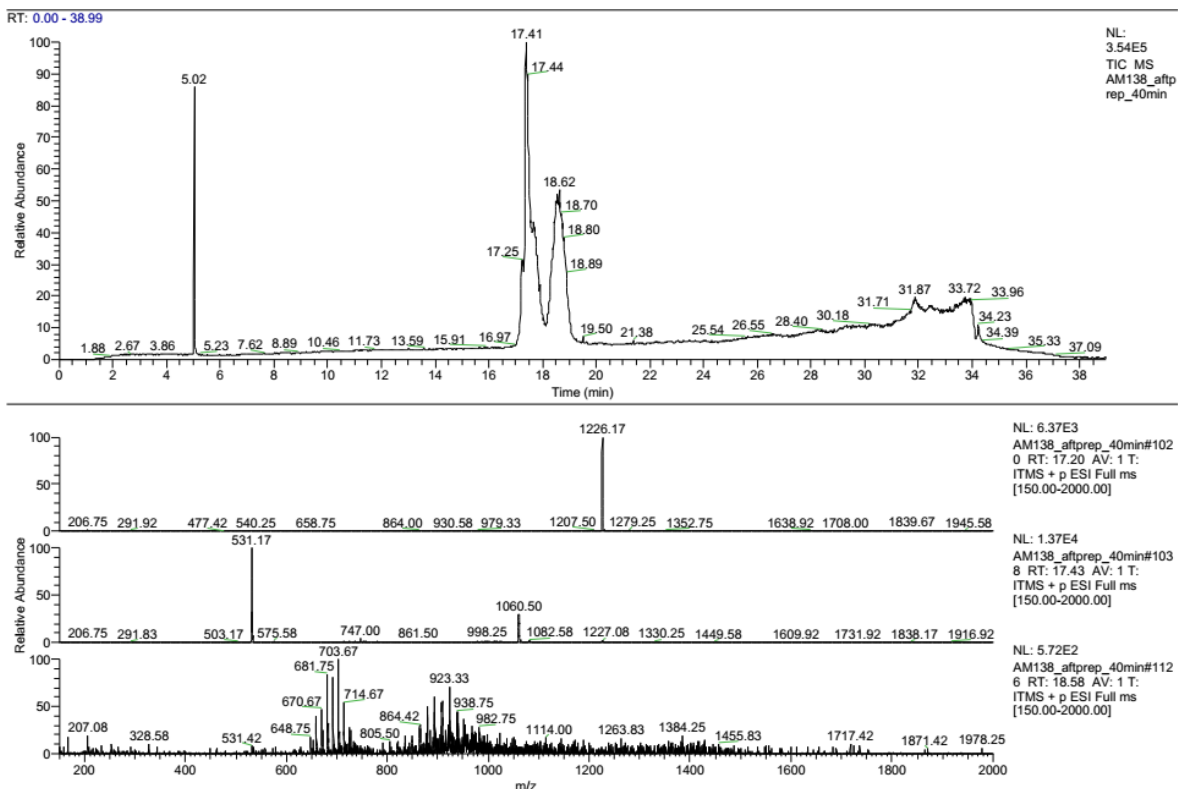


Figure 152 LCMS analysis of purified compound 31(in the synthesis of compound 26 with reverse order of loops attachment), which could not be separated from TBTA.

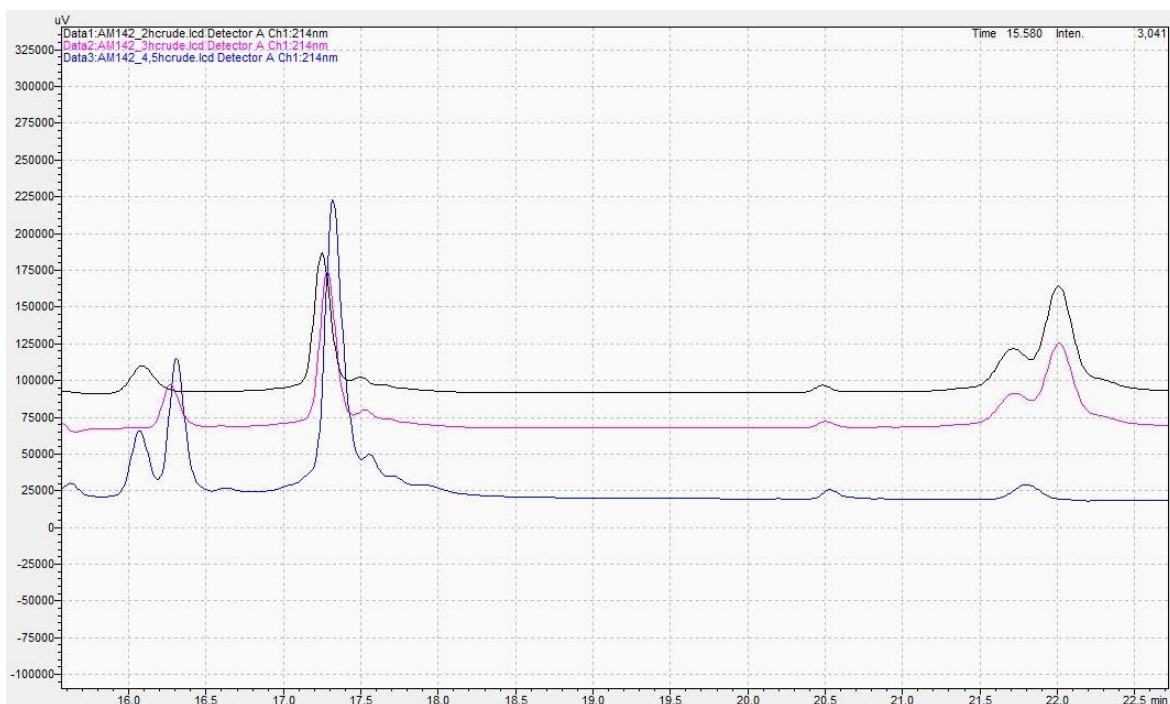


Figure 153 Progress of TIPS removal from compound 31 to obtain compound 32 in the synthesis of 26 with reverse order of loops attachment as analysed by analytical HPLC.

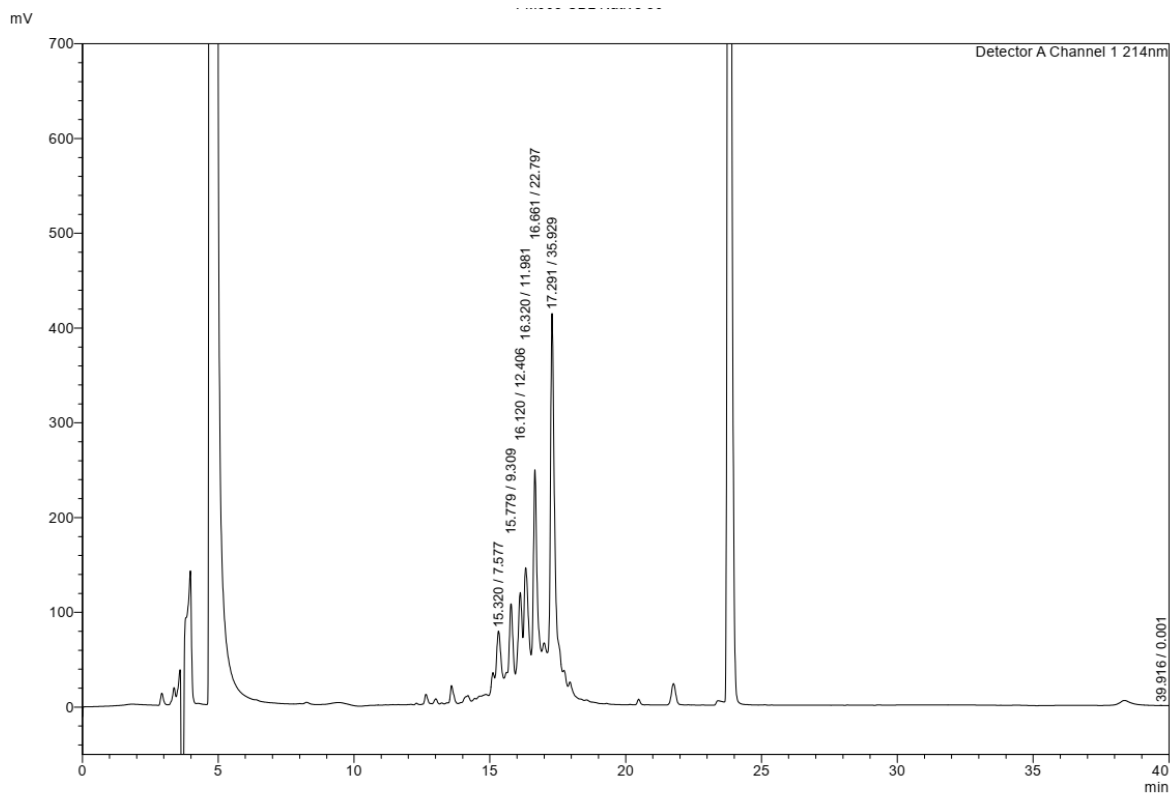


Figure 154 analytical HPLC of purified compound 32 (in the synthesis of compound 26 with reverse order of loops attachment as analysed by analytical HPLC).

E:\LCMS\...AM142\_2hcrude\_40min

23/05/2018 10:52:27

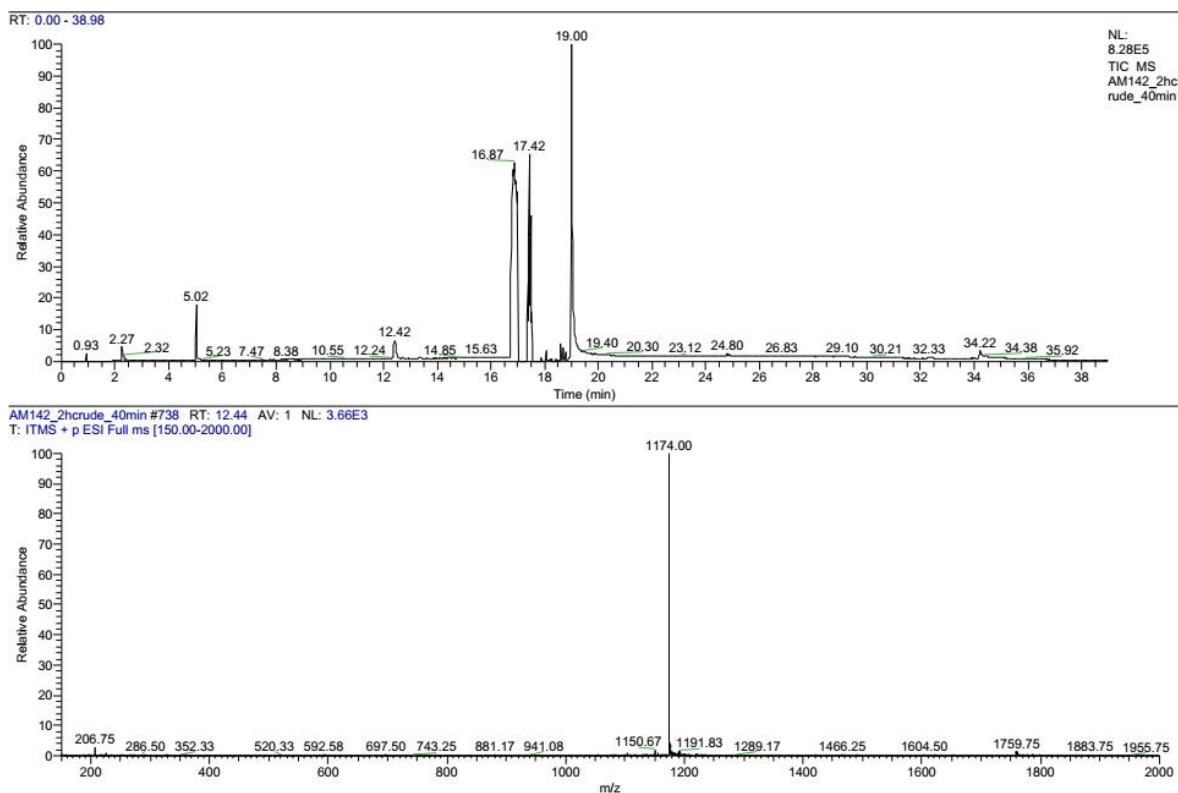
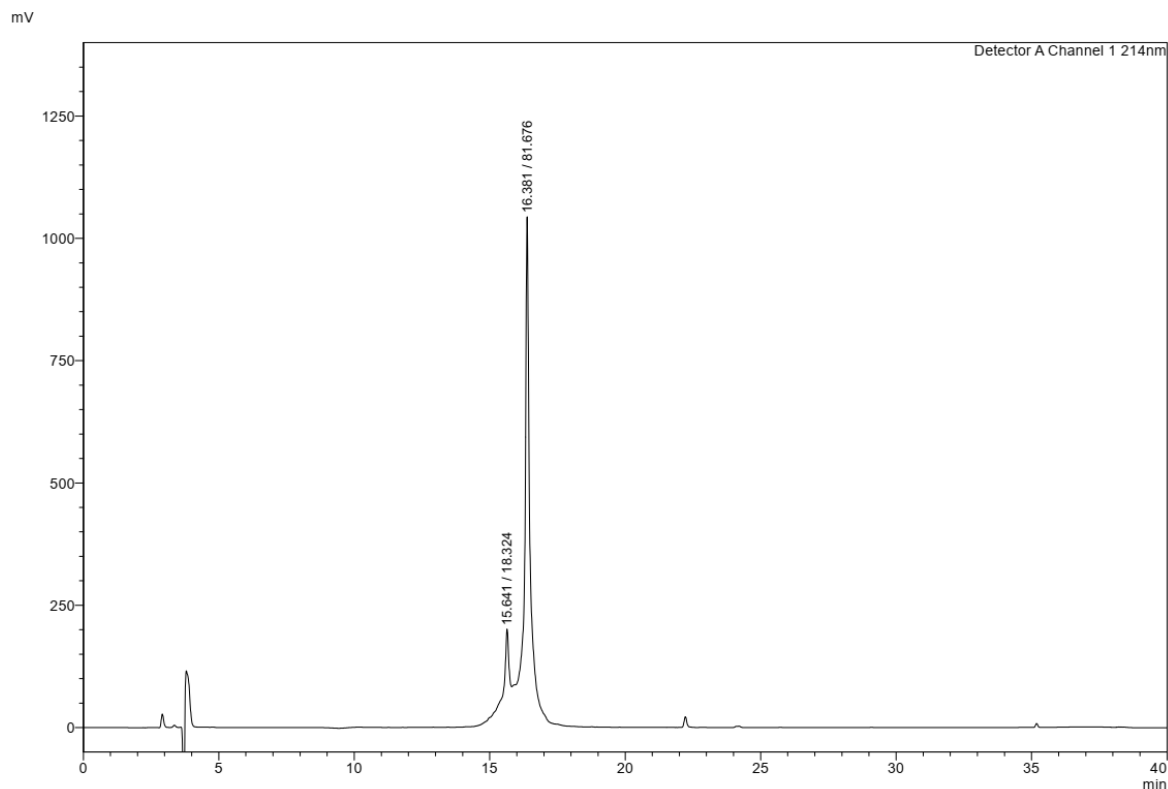
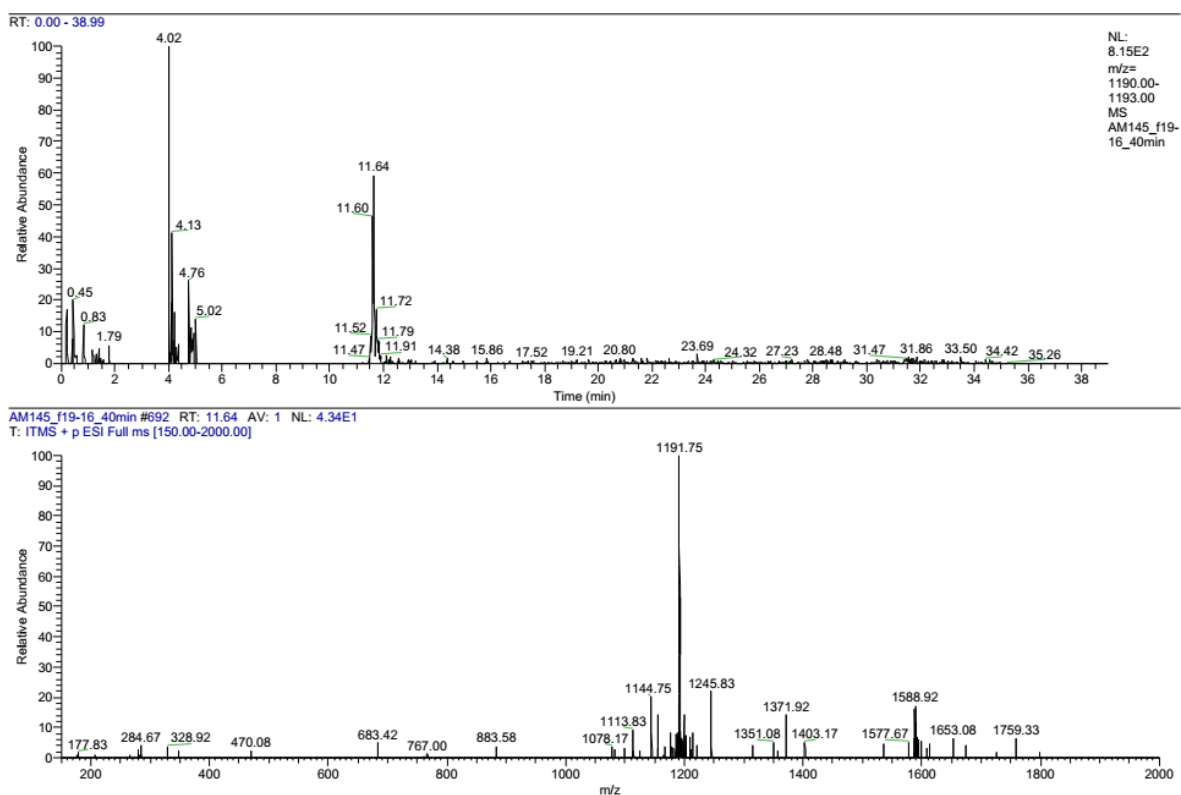


Figure 155 LCMS analysis of TIPS removal from compound 31 to obtain compound 32 in the synthesis of compound 26 with reverse order of loops attachment. Crude reaction mixture after 2h.



E:\LCMS\...AM145\_f19-16\_40min

14/07/2018 21:14:28



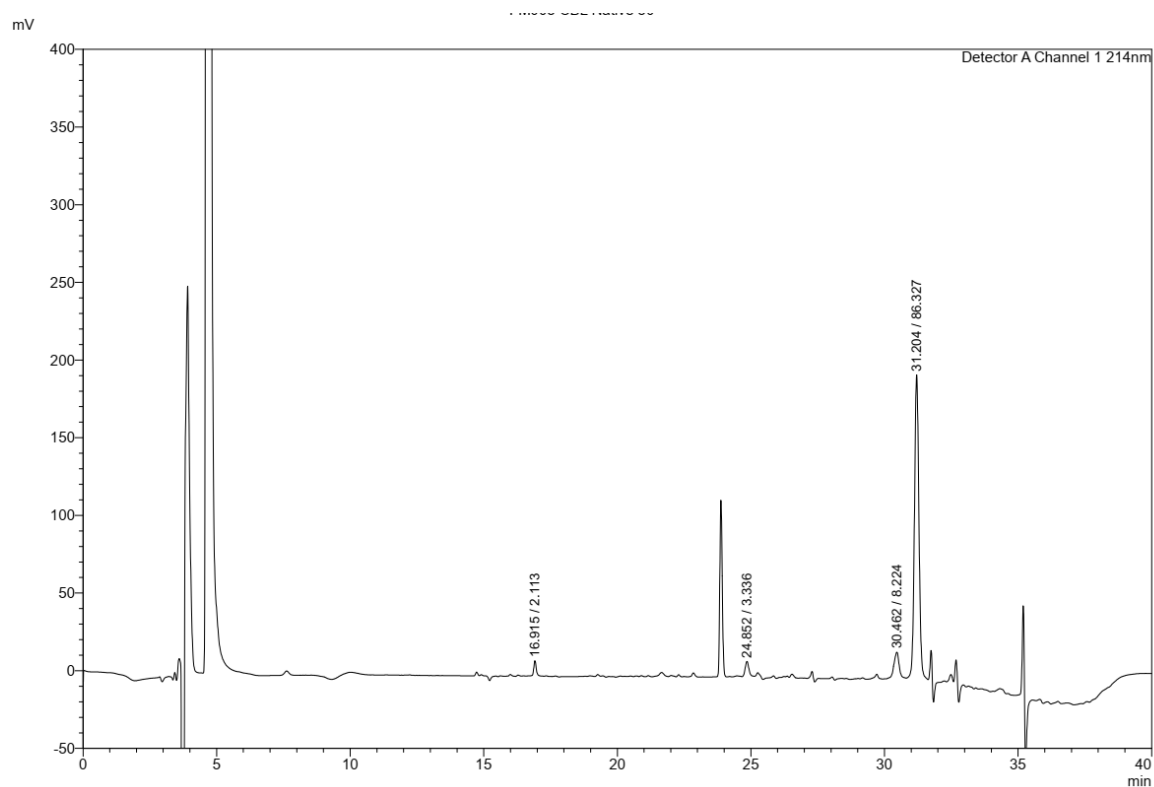


Figure 158 analytical HPLC: attachment of loop 2 to TAC scaffold in an attempt of one-pot synthesis of gp120 mimic.

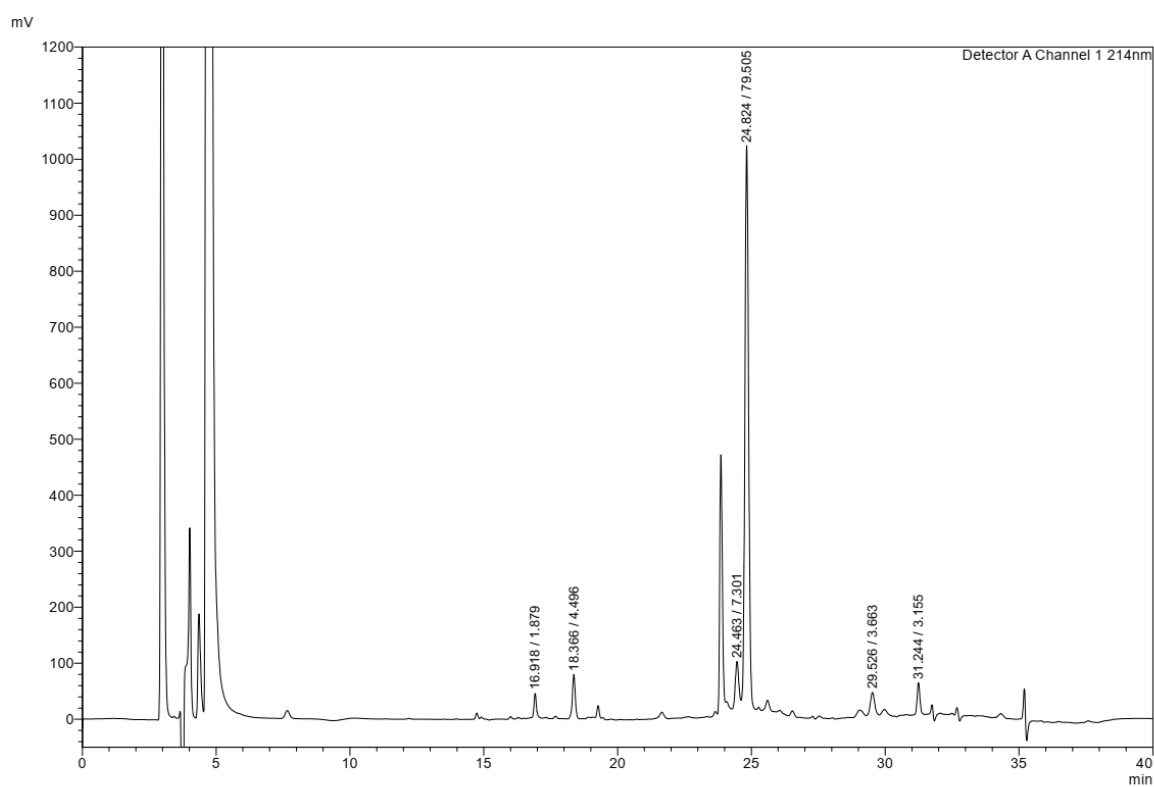


Figure 159 analytical HPLC of TES deprotection to obtain compound 23 in an attempt of one-pot synthesis of gp120 mimic.

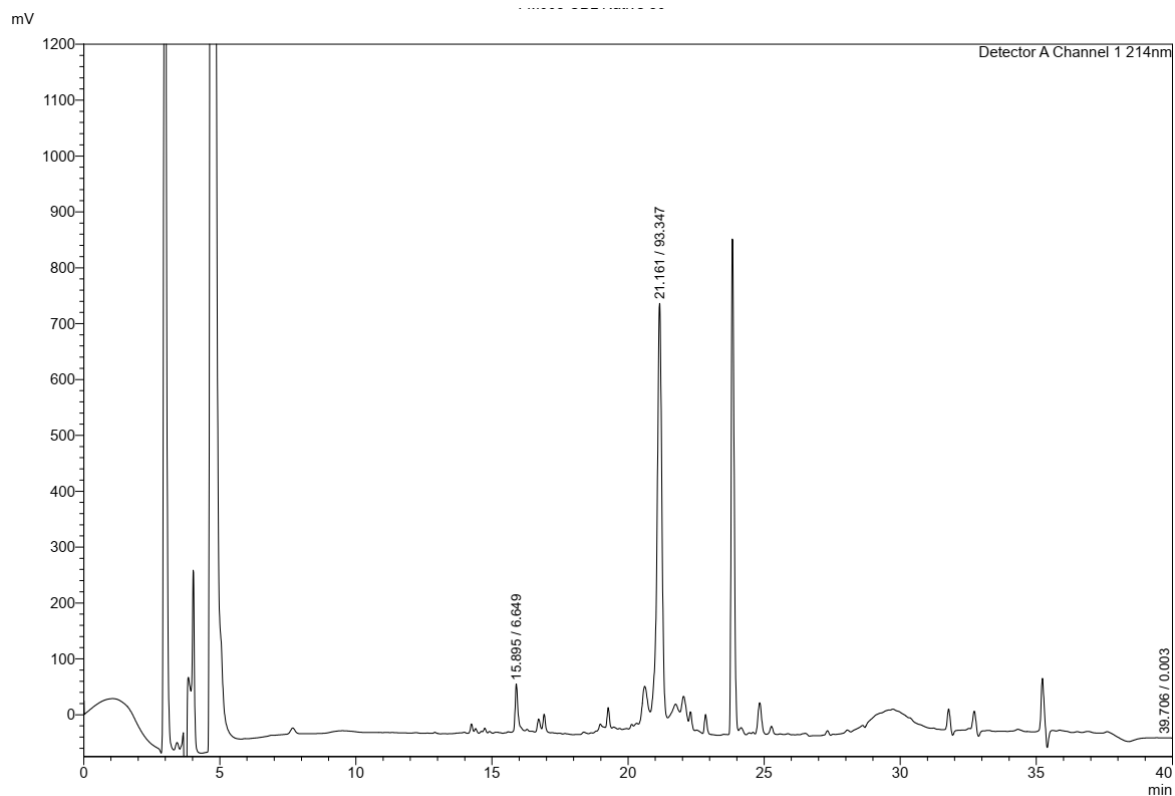


Figure 160 analytical HPLC of loop 1 attachment in the synthesis of compound 24 in an attempt of one-pot synthesis of gp120 mimic.

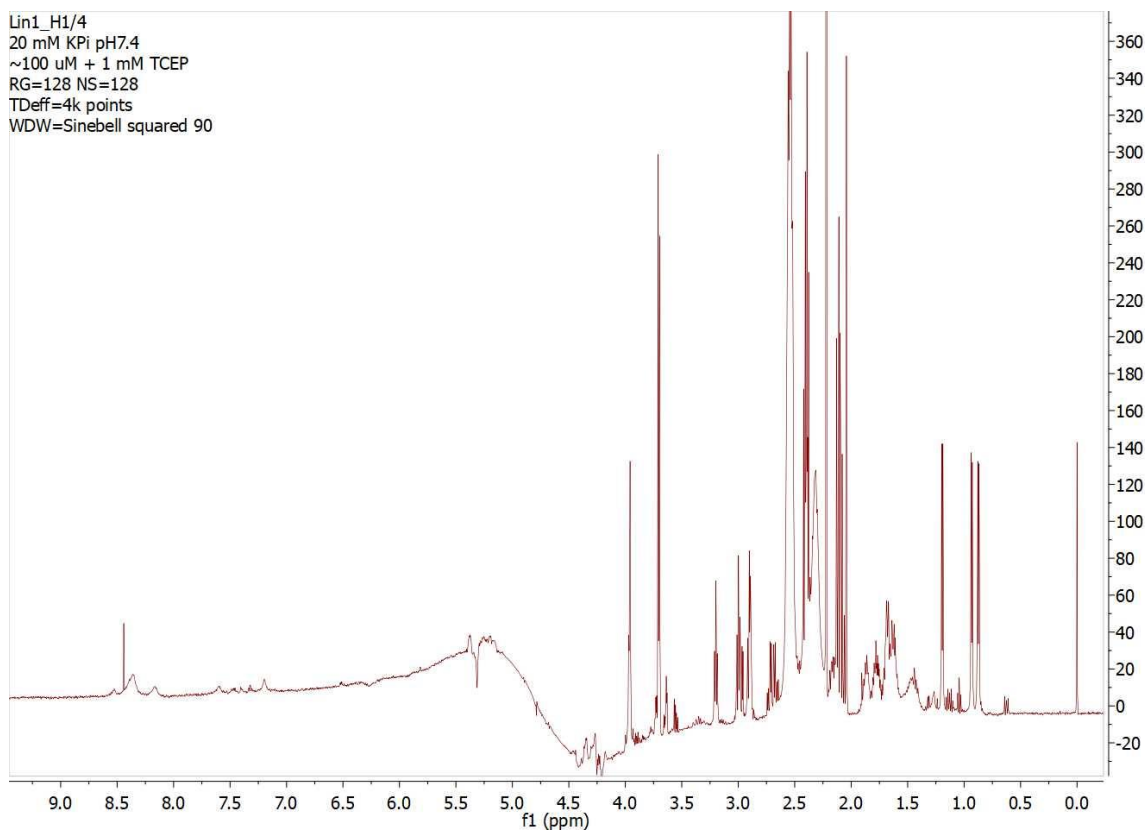
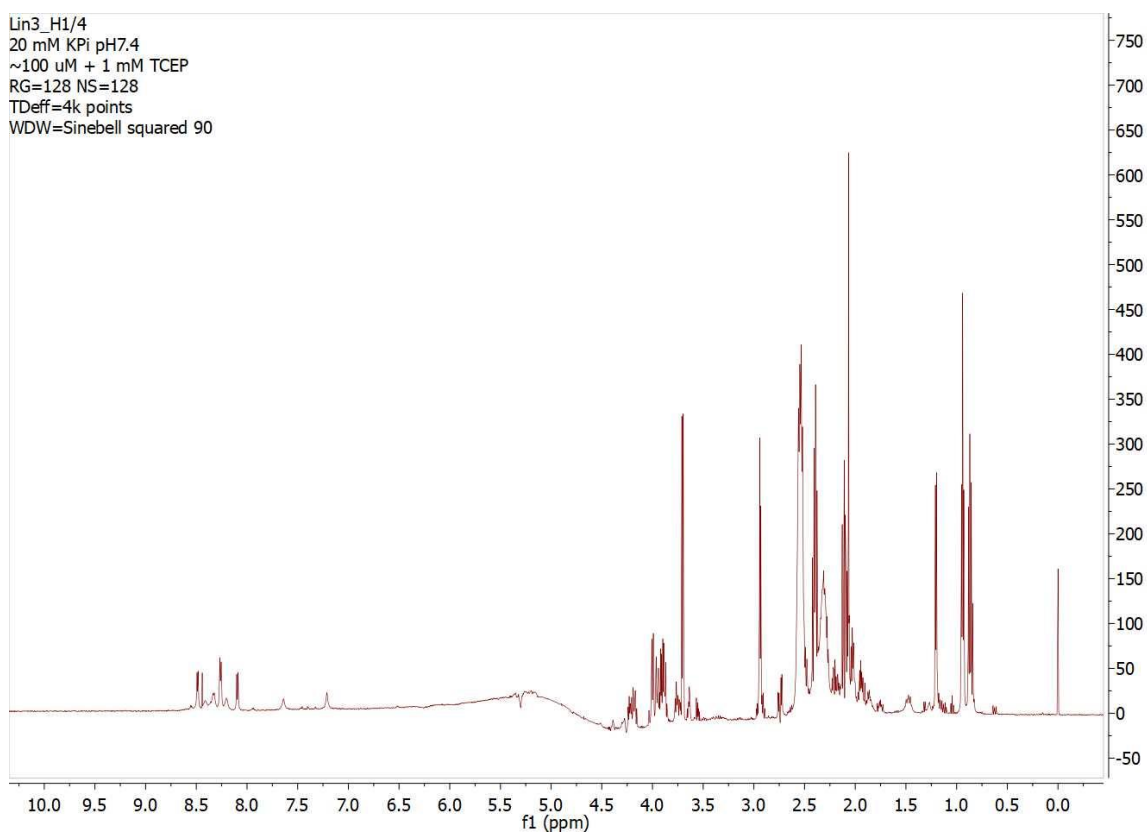
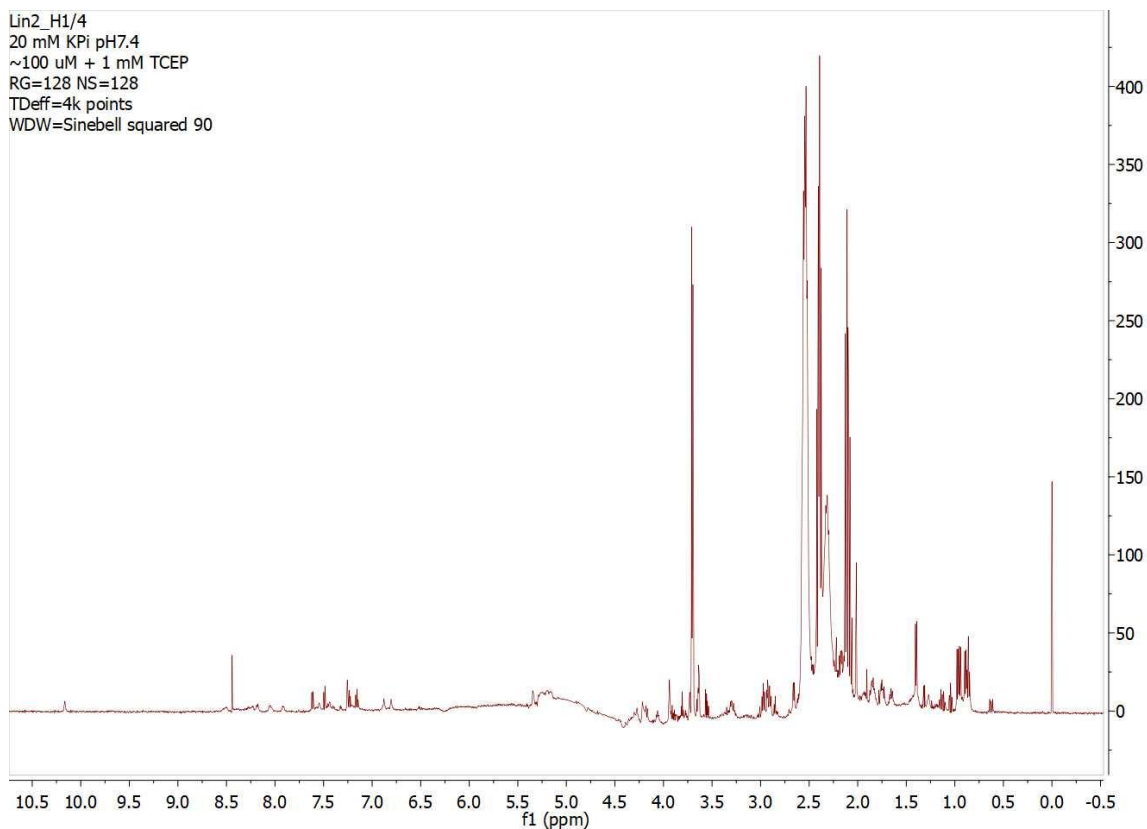


Figure 161 <sup>1</sup>H-NMR analysis of linear peptide 1 with added TCEP. Water signal was suppressed.





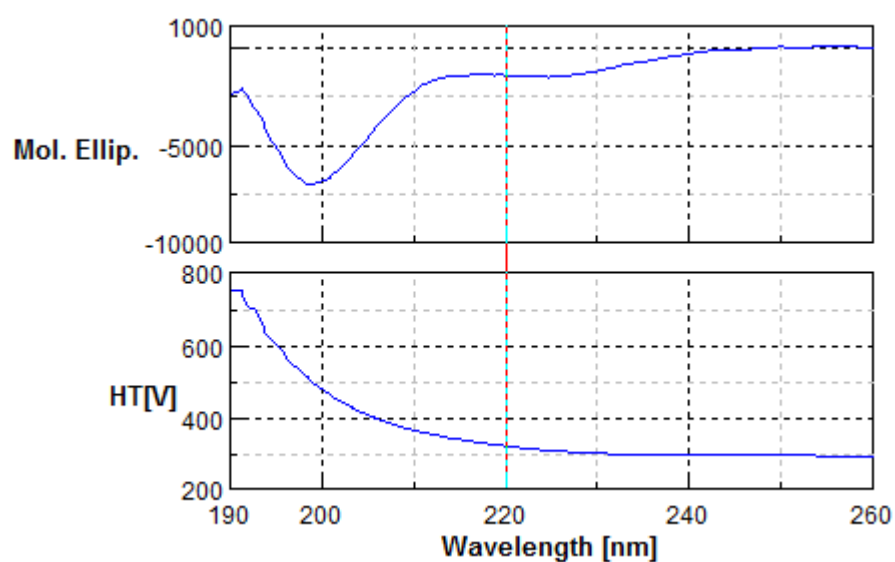


Figure 164 CD spectrum of linear peptide 1 with TCEP.

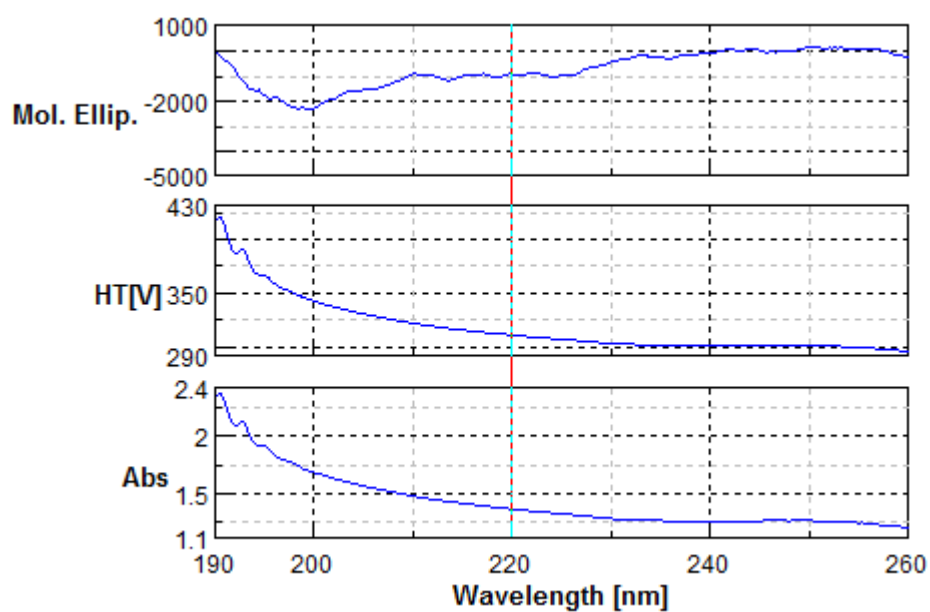


Figure 165 CD spectrum of linear peptide 2 with TCEP.

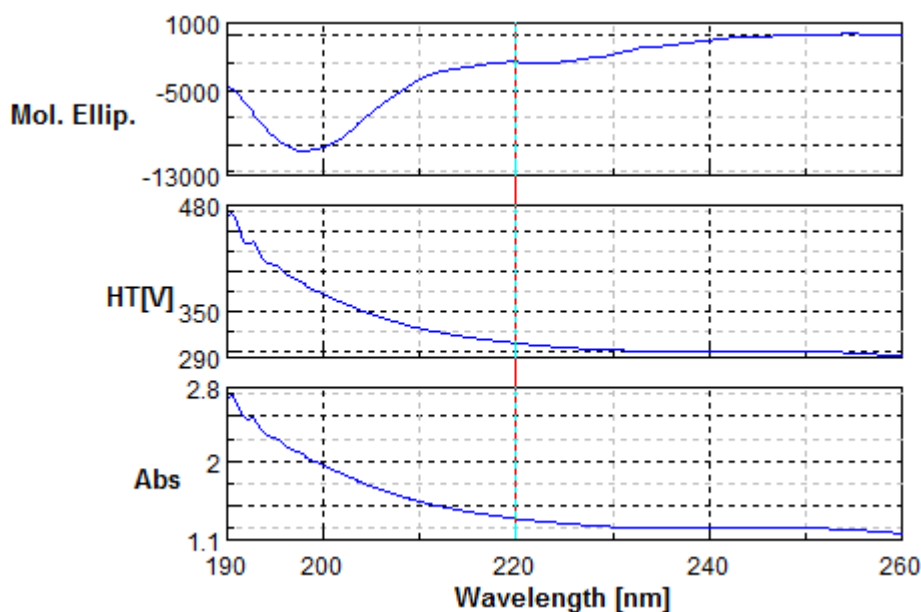


Figure 166 CD spectrum of linear peptide 3 with TCEP.

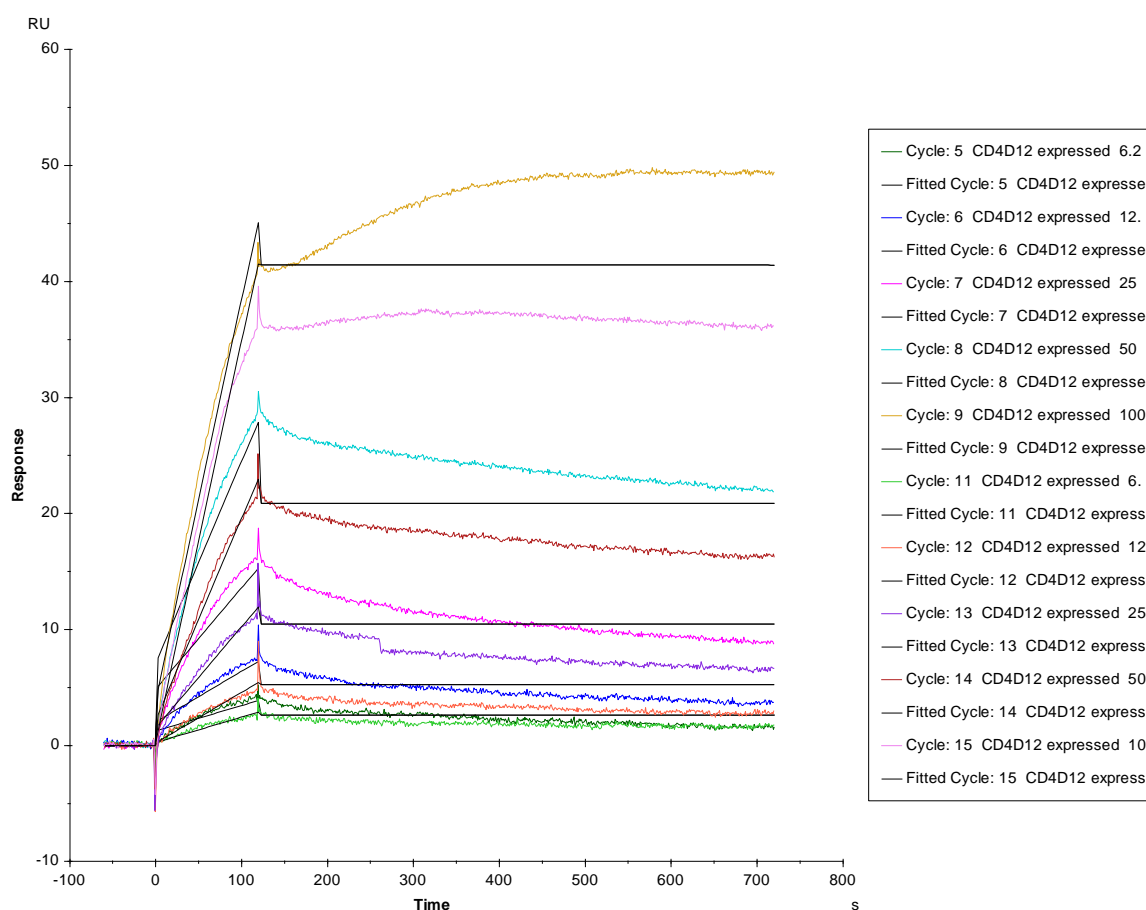


Figure 167 Replicate 2 of interaction of analyte CD4D12<sub>ex</sub> with immobilised ligand gp120 at 25C. The coloured lines depict the double-referenced sensograms obtained from duplicate injections of 100, 50, 25, 12.5, 6.3 nM analyte CD4D12<sub>ex</sub> across 1000 RU immobilised ligand gp120. The ligand surface was regenerated with a 60 s injection of 10 mM glycine pH 1.75 between each binding cycle. The black lines depict the global fit of the data to a 1:1 Langmuir interaction model, yielding  $k_a = 2.69 \text{ M}^{-1} \cdot \text{s}^{-1}$ ,  $k_d = 1.22 \cdot 10^{-6} \cdot \text{s}^{-1}$ ,  $K_d = 4.53 \cdot 10^{-7} \text{ M}$ .

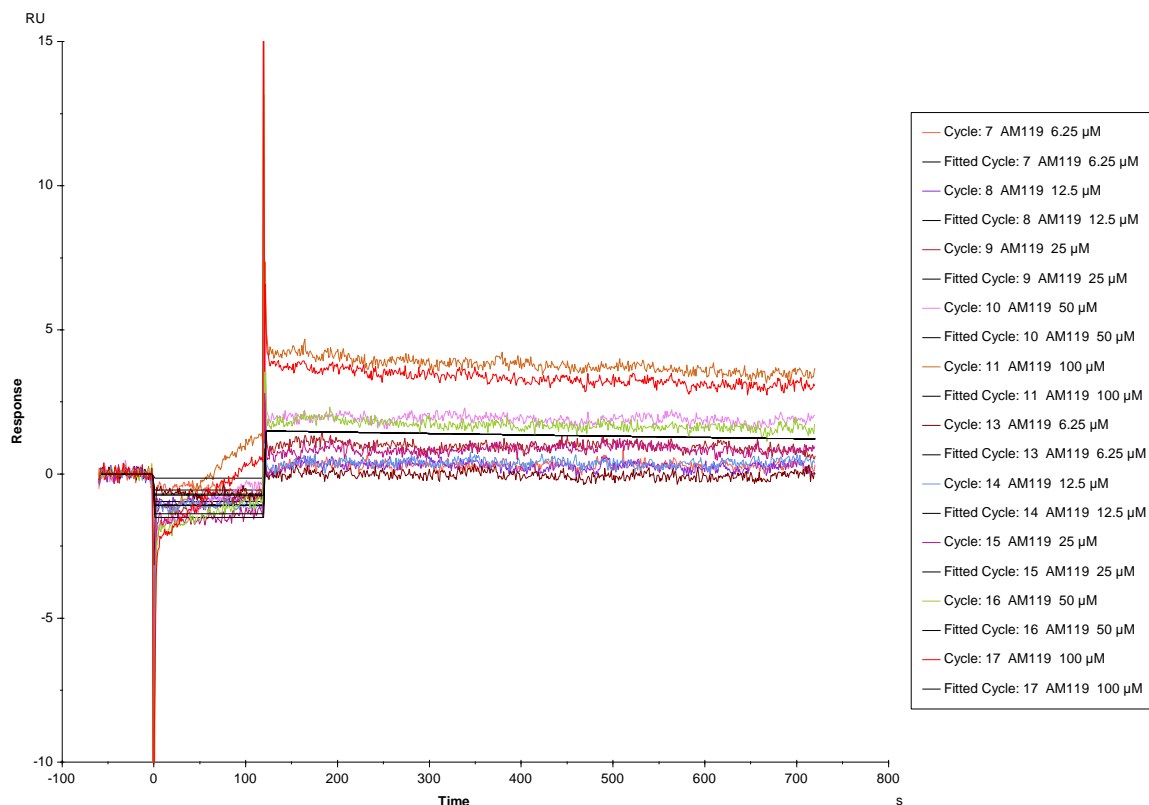


Figure 168 Replicate 2 of performed experiment to measure binding of mimic 22 to the immobilised CD4D12 at 25C.

The coloured lines depict the double-referenced sensograms obtained from duplicate injections of 100, 50, 25, 12.5, 6.25  $\mu\text{M}$  analyte 22 across 1200 RU immobilised ligand CD4D12. The ligand surface was regenerated with two 30 s injections of 10 mM glycine pH 1.75 between each binding cycle. The black lines depict the global fit of the data to a 1:1 Langmuir interaction model, yielding  $k_a = 2.04 \cdot 10^9 \text{ M}^{-1} \cdot \text{s}^{-1}$ ,  $k_d = 3.45 \cdot 10^{-4} \cdot \text{s}^{-1}$ ,  $K_d = 1.69 \cdot 10^{-13} \text{ M}$ .

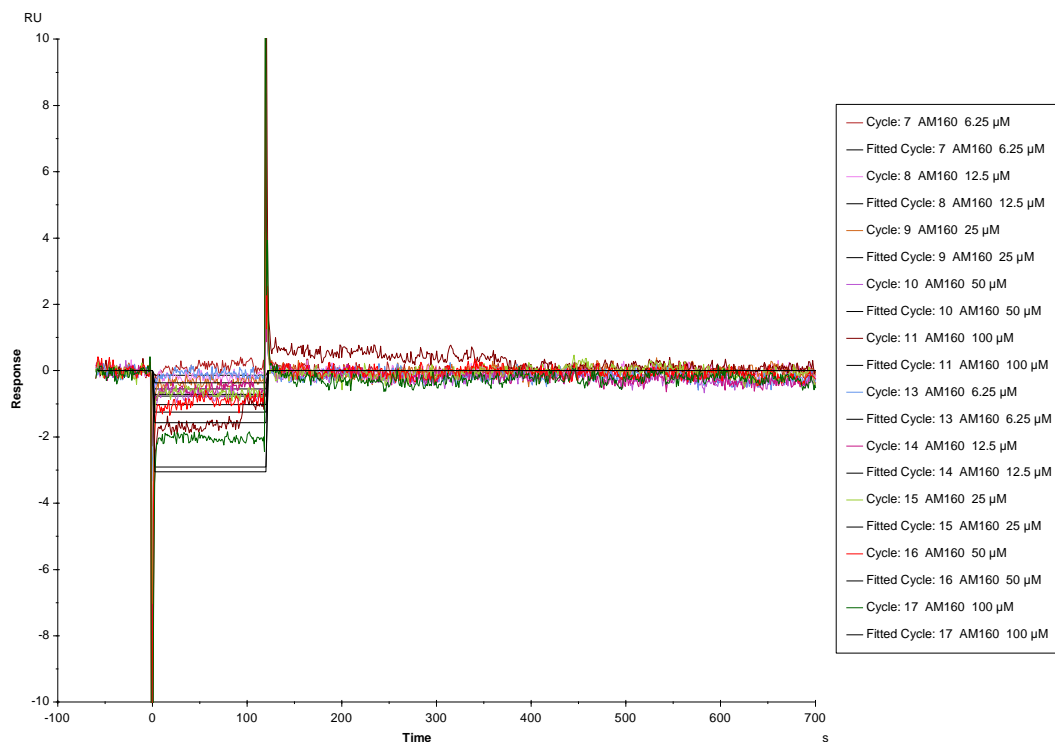


Figure 169 Replicate 2 of performed experiment to measure binding of mimic 26 to the immobilised CD4D12 at 25C.

The coloured lines depict the double-referenced sensograms obtained from duplicate injections of 100, 50, 25, 12.5, 6.25  $\mu$ M analyte 26 across 1200 RU immobilised ligand CD4D12. The ligand surface was regenerated with two 30 s injections of 10 mM glycine pH 1.75 between each binding cycle. The black lines depict the global fit of the data to a 1:1 Langmuir interaction model.

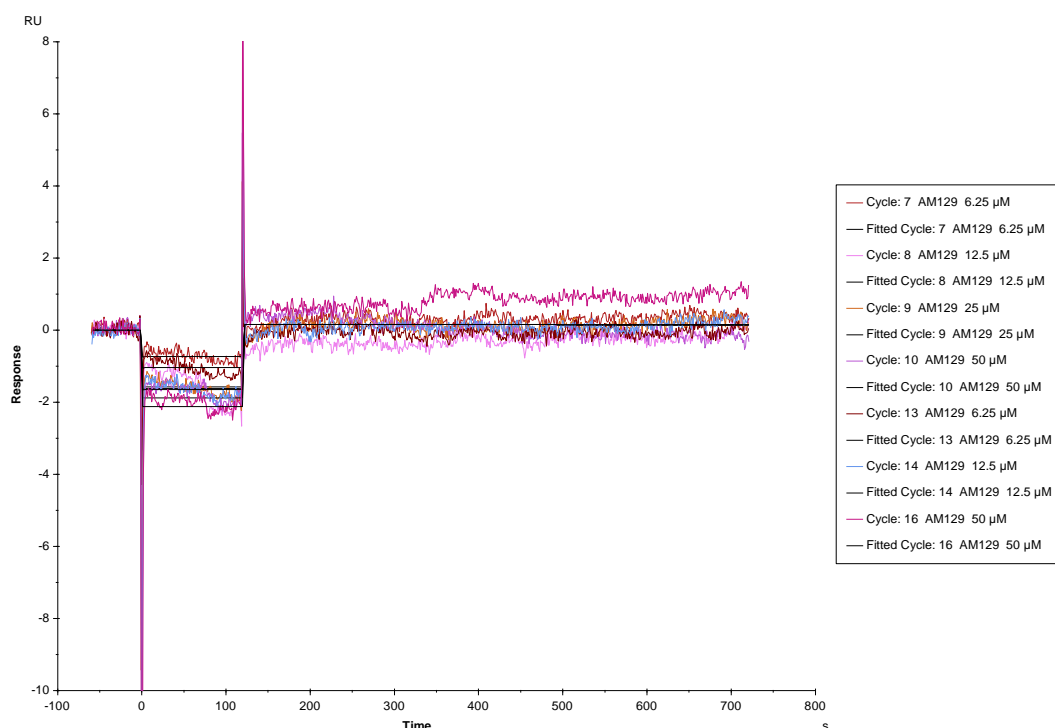


Figure 170 Replicate 2 of performed experiment to measure binding of mimic 30 to the immobilised CD4D12 at 25C.

The coloured lines depict the double-referenced sensograms obtained from duplicate injections of 100, 50, 25, 12.5, 6.25  $\mu$ M analyte 30 across 1200 RU immobilised ligand CD4D12. The ligand surface was regenerated with two 30 s injections of 10 mM glycine pH 1.75 between each binding cycle. The black lines depict the global fit of the data to a 1:1 Langmuir interaction model.

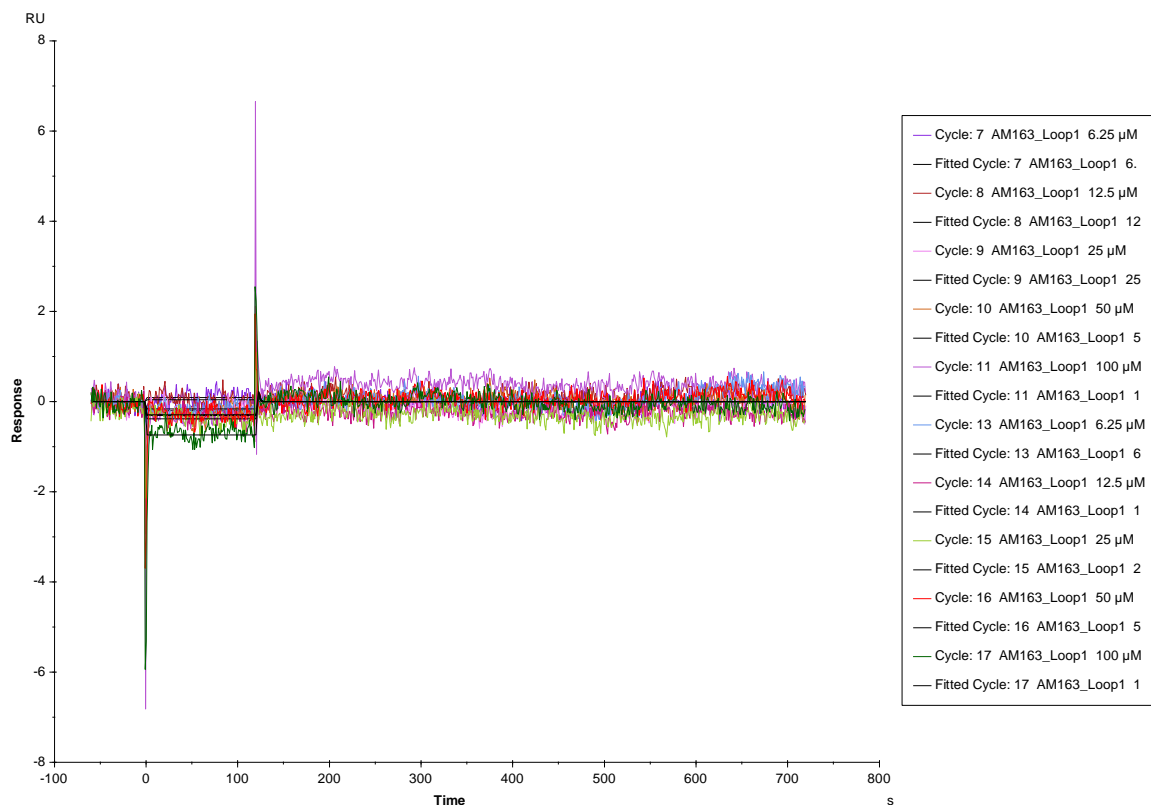


Figure 171 Replicate 2 of performed experiment to measure binding of loop 1 to the immobilised CD4D12 at 25C.

The coloured lines depict the double-referenced sensograms obtained from duplicate injections of 100, 50, 25, 12.5, 6.25  $\mu$ M analyte loop 1 across 1200 RU immobilised ligand CD4D12. The ligand surface was regenerated with two 30 s injections of 10 mM glycine pH 1.75 between each binding cycle. The black lines depict the global fit of the data to a 1:1 Langmuir interaction model.

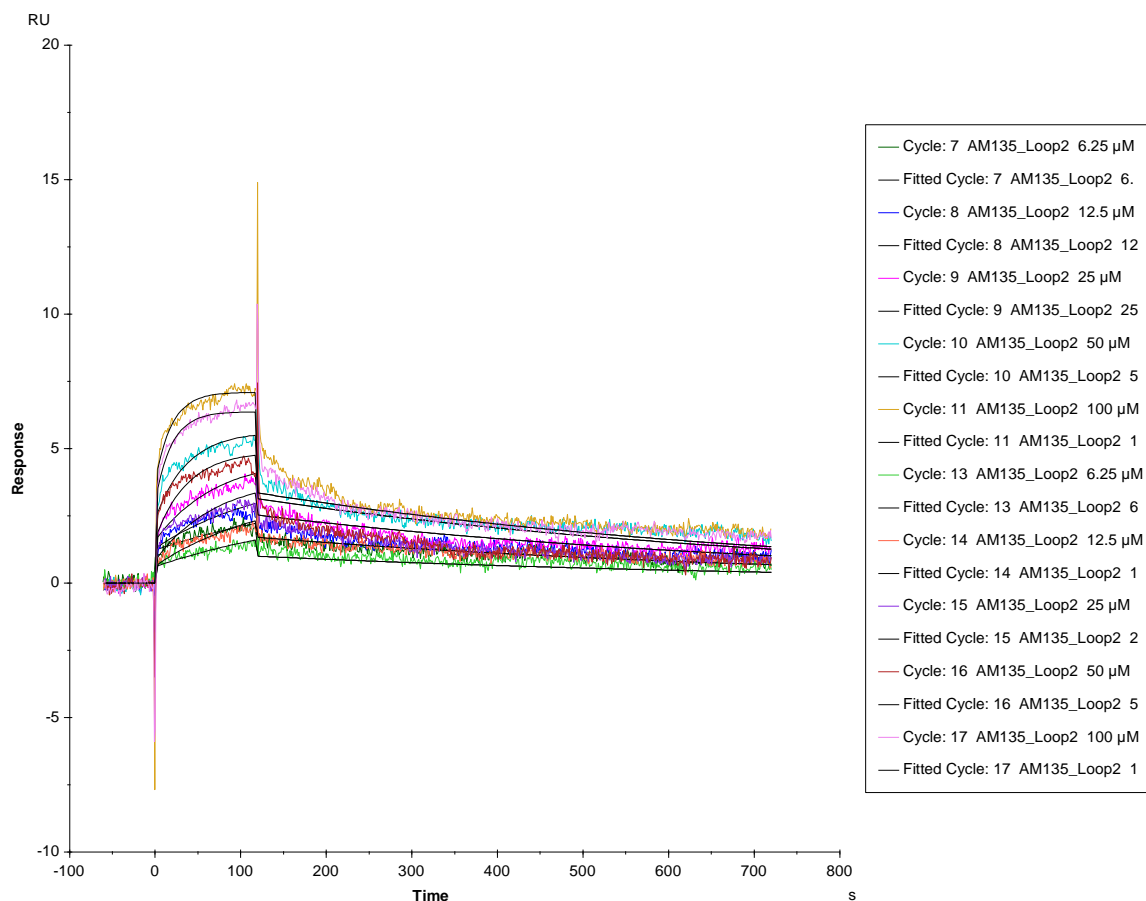


Figure 172 Replicate 2 of performed experiment to measure binding of loop 2 to the immobilised CD4D12 at 25C.

The coloured lines depict the double-referenced sensograms obtained from duplicate injections of 100, 50, 25, 12.5, 6.25  $\mu\text{M}$  analyte loop 2 across 1200 RU immobilised ligand CD4D12. The ligand surface was regenerated with two 30 s injections of 10 mM glycine pH 1.75 between each binding cycle. The black lines depict the global fit of the data to a 1:1 Langmuir interaction model, yielding  $k_a = 507.0 \text{ M}^{-1} \cdot \text{s}^{-1}$ ,  $k_d = 1.508 \cdot 10^{-3} \text{ s}^{-1}$ ,  $K_d = 2.974 \cdot 10^{-6} \text{ M}$ .

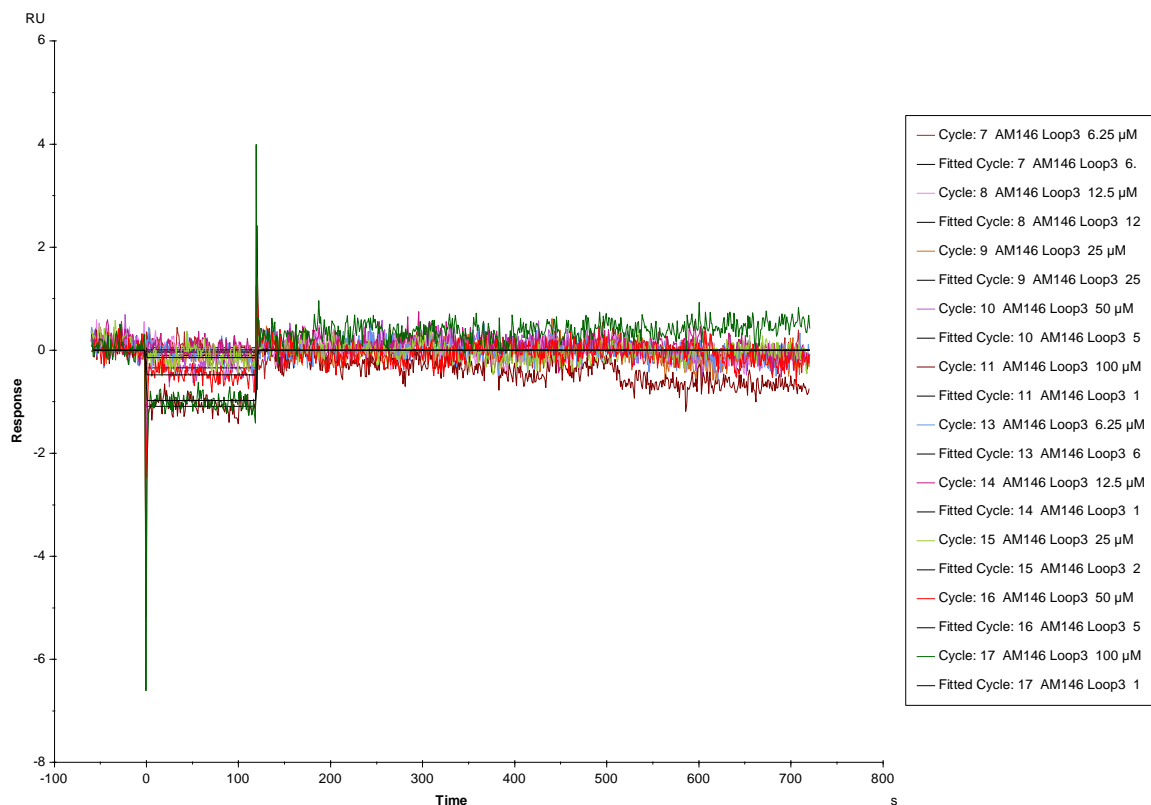


Figure 173 Replicate 2 of performed experiment to measure binding of loop 3 to the immobilised CD4D12 at 25C.

The coloured lines depict the double-referenced sensograms obtained from duplicate injections of 100, 50, 25, 12.5, 6.25  $\mu$ M analyte loop 3 across 1200 RU immobilised ligand CD4D12. The ligand surface was regenerated with two 30 s injections of 10 mM glycine pH 1.75 between each binding cycle. The black lines depict the global fit of the data to a 1:1 Langmuir interaction model.

## Bibliography

- [1] D. A. Henderson, *Vaccine* **2011**, 29, D7-D9.
- [2] B. Greenwood, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2014**, 369, 20130433.
- [3] WHO, *Ten Years Public Heal. 2007-2017* **2017**, 6.
- [4] WHO, *Who* **2013**.
- [5] J. Wallis, D. P. Shenton, R. C. Carlisle, *Clin. Exp. Immunol.* **2019**, cei.13287.
- [6] J. C. Becerra, L. S. Bildstein, J. S. Gach, *Microb. Cell* **2016**, 3, 450-474.
- [7] C. Apetrei, B. Gormus, I. Pandrea, M. Metzger, P. Haaft, L. N. Martin, R. Bohm, X. Alvarez, G. Koopman, M. Murphey-corb, et al., **2004**, 78, 11506-11518.
- [8] UNAIDS Communications and Global Advocacy, **2018**, 1-5.
- [9] C. A. Didigu, R. W. Doms, *Viruses* **2012**, 4, 309-324.
- [10] C. B. Wilen, J. C. Tilton, R. W. Doms, B. Walker, A. Mcmichael, C. B. Wilen, J. C. Tilton, R. W. Doms, R. Craigie, F. D. Bushman, et al., *Cold Spring Harb. Perspect. Med.* **2012**, 2, 1-14.
- [11] P. Poignard, E. O. Saphire, P. W. Parren, D. R. Burton, *Annu. Rev. Immunol.* **2001**, 19, 253-274.
- [12] M. A. Checkley, B. G. Luttge, E. O. Freed, *J. Mol. Biol.* **2011**, 410, 582-608.
- [13] H. B. Bernstein, S. P. Tucker, E. Hunter, J. S. Schutzbach, R. W. Compans, *J. Virol.* **1994**, 68, 463-8.



- [14] K. J. Doores, C. Bonomelli, D. J. Harvey, S. Vasiljevic, R. A. Dwek, D. R. Burton, M. Crispin, C. N. Scanlan, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 13800-5.
- [15] R. Pantophlet, D. R. Burton, *Annu. Rev. Immunol.* **2006**, *24*, 739-769.
- [16] P. D. Kwong, R. Wyatt, J. Robinson, R. W. Sweet, J. Sodroski, W. a Hendrickson, *Nature* **1998**, *393*, 648-659.
- [17] P. Acharya, S. Lusvarghi, C. A. Bewley, P. D. Kwong, *Expert Opin. Ther. Targets* **2015**, *19*, 765-783.
- [18] A. F. Labrijn, P. Poignard, A. Raja, B. Michael, K. Delgado, M. Franti, J. Binley, V. Vivona, C. Grundner, C. Huang, et al., **2003**, *77*, 10557-10565.
- [19] J. M. Binley, Y. E. A. Ban, E. T. Crooks, D. Eggink, K. Osawa, W. R. Schief, R. W. Sanders, *J. Virol.* **2010**, *84*, 5637-5655.
- [20] P. L. Moore, E. S. Gray, C. K. Wibmer, J. N. Bhiman, M. Nonyane, D. J. Sheward, T. Hermanus, S. Bajimaya, N. L. Tumba, R. Abrahams, et al., *Nat Med* **2012**, *18*, 1688-1692.
- [21] G. Frey, H. Peng, S. Rits-Volloch, M. Morelli, Y. Cheng, B. Chen, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 3739-44.
- [22] P. Poignard, M. Moulard, E. Golez, V. Vivona, M. Franti, S. Venturini, M. Wang, P. W. H. I. Parren, D. R. Burton, *J. Virol.* **2003**, *77*, 353-65.
- [23] Moore PL, Crooks ET, Porter L, Zhu P, Cayanan CS, Grise H, Corcoran P, Zwick MB, Franti M, Morris L, et al., *J. Virol.* **2006**, *80*, 2515-2528.
- [24] P. Zhu, J. Liu, J. Bess, E. Chertova, J. D. Lifson, H. Grisé, G. A. Ofek, K. A. Taylor, K. H. Roux, *Nature* **2006**, *441*, 847-852.
- [25] J. S. Klein, P. N. P. Gnanapragasam, R. P. Galimidi, C. P. Foglesong, A. P. West, P. J. Bjorkman, *Proc. Natl. Acad. Sci.* **2009**, *106*, 7385-7390.

- [26] P. D. Kwong, M. L. Doyle, D. J. Casper, C. Cicala, S. A. Leavitt, S. Majeed, T. D. Steenbeke, M. Venturi, I. Chaiken, M. Fung, et al., *Nature* **2002**, *420*, 678-682.
- [27] H. Mouquet, *Trends Immunol.* **2014**, *35*, 549-561.
- [28] M. H. V. Van Regenmortel, *Front. Immunol.* **2012**, *3*, 1-19.
- [29] S. Paul, S. Planque, Y. Nishiyama, M. Escobar, Hanson Carl, **2011**, *9*, 1027-1043.
- [30] R. Pantophlet, I. A. Wilson, D. R. Burton, *J. Virol.* **2003**, *77*, 5889-901.
- [31] R. Pantophlet, E. Ollmann Saphire, P. Poignard, P. W. H. I. Parren, I. A. Wilson, D. R. Burton, *J. Virol.* **2003**, *77*, 642-58.
- [32] B. Dey, M. Pancera, K. Svehla, Y. Shu, S.-H. Xiang, J. Vainshtein, Y. Li, J. Sodroski, P. D. Kwong, J. R. Mascola, et al., *J. Virol.* **2007**, *81*, 5579-5593.
- [33] R. Pantophlet, D. R. Burton, *Trends Mol. Med.* **2003**, *9*, 468-473.
- [34] S. A. Kalams, S. D. Parker, M. Elizaga, B. Metch, S. Edupuganti, J. Hural, S. De Rosa, D. K. Carter, K. Rybczyk, I. Frank, et al., *J. Infect. Dis.* **2013**, *208*, 818-829.
- [35] F. García, N. Climent, A. C. Guardo, C. Gil, A. León, B. Autran, J. D. Lifson, J. Martínez-Picado, J. Dalmau, B. Clotet, et al., *Sci. Transl. Med.* **2013**, *5*, DOI 10.1126/scitranslmed.3004682.
- [36] A. Lind, M. Sommerfelt, J. O. Holmberg, I. Baksaas, B. Sørensen, D. Kvale, *Scand. J. Infect. Dis.* **2012**, *44*, 566-572.
- [37] J. H. Kim, P. Kunasol, N. L. Michael, M. L. Robb, P. Thongcharoen, C. Khamboonruang, *N. Engl. J. Med.* **2009**, *361*, 2209-2220.
- [38] J. L. Excler, J. Ake, M. L. Robb, J. H. Kim, S. A. Plotkin, *Clin. Vaccine Immunol.* **2014**, *21*, 1023-1036.

- [39] A. Saez-Cirion, B. Jacquelin, F. Barré-Sinoussi, M. Müller-Trutwin, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2014**, *369*, 20130436.
- [40] J. de Las Rivas, C. Fontanillo, *PLoS Comput. Biol.* **2010**, *6*, 1-8.
- [41] E. Cukuroglu, H. B. Engin, A. Gursoy, O. Keskin, *Prog. Biophys. Mol. Biol.* **2014**, *116*, 165-173.
- [42] J. A. Robinson, S. DeMarco, F. Gombert, K. Moehle, D. Obrecht, *Drug Discov. Today* **2008**, *13*, 944-951.
- [43] P. Harjes, E. E. Wanker, *Trends Biochem. Sci.* **2003**, *28*, 425-433.
- [44] H. zur Hausen, *Nat. Rev. Cancer* **2002**, *2*, 342-350.
- [45] A. A. Bogan, K. S. Thorn, *J. Mol. Biol.* **1998**, *280*, 1-9.
- [46] J. Janin, *Proteins Struct. Funct. Genet.* **1995**, *21*, 30-39.
- [47] A. A. Bogan, K. S. Thorn, *J. Mol. Biol.* **1998**, *280*, 1-9.
- [48] J. a. Robinson, S. DeMarco, F. Gombert, K. Moehle, D. Obrecht, *Drug Discov. Today* **2008**, *13*, 944-951.
- [49] M. R. Arkin, Y. Tang, J. A. Wells, *Chem Biol* **2014**, *21*, 1102-1114.
- [50] H. Yin, A. D. Hamilton, *Angew. Chemie - Int. Ed.* **2005**, *44*, 4130-4163.
- [51] Y. Che, A. M. Gilbert, V. Shanmugasundaram, M. C. Noe, *Bioorg. Med. Chem. Lett.* **2018**, 1-8.
- [52] A. D. Cunningham, N. Qvit, D. Mochly-Rosen, *Curr. Opin. Struct. Biol.* **2017**, *44*, 59-66.
- [53] O. Karni-Schmidt, M. Lokshin, C. Prives, *Annu. Rev. Pathol. Mech. Dis.* **2016**, *11*, 617-644.

- [54] A. Burgess, K. M. Chia, S. Haupt, D. Thomas, Y. Haupt, E. Lim, *Front. Oncol.* **2016**, *6*, 1-7.
- [55] I. M. A. Nooren, J. M. Thornton, *EMBO J.* **2003**, *22*, 3486-3492.
- [56] D. E. Scott, A. R. Bayly, C. Abell, J. Skidmore, *Nat. Rev. Drug Discov.* **2016**, *15*, 533-550.
- [57] J. R. Bradford, J. A. Siepen, and D. R. Westhead, *Fundamentals of Protein Structure and Function*, **2013**.
- [58] G. D. Hartman, M. S. Egbertson, W. Halczenko, W. L. Laswell, M. E. Duggan, R. L. Smith, A. M. Naylor, P. D. Manno, R. J. Lynch, G. Zhang, et al., *J. Med. Chem.* **1992**, *35*, 4640-4642.
- [59] J. Morinière, S. Rousseaux, U. Steuerwald, M. Soler-López, S. Curtet, A. L. Vitte, J. Govin, J. Gaucher, K. Sadoul, D. J. Hart, et al., *Nature* **2009**, *461*, 664-668.
- [60] P. H. Kussie, S. Gorina, V. Marechal, B. Elenbaas, J. Moreau, A. J. Levine, N. P. Pavletich, *Science (80-. )*. **1996**, *274*, 948-953.
- [61] M. Rickert, X. Wang, M. J. Boulanger, N. Goriatcheva, K. C. Garcia, *Science (80-. )*. **2005**, *308*, 1477-1480.
- [62] L. Nevola, E. Giralt, *Chem. Commun.* **2015**, *51*, 3302-3315.
- [63] O. Byron, B. Vestergaard, **2015**, 76-86.
- [64] M. Zhou, Q. Li, R. Wang, *ChemMedChem* **2016**, *11*, 738-756.
- [65] X. Du, Y. Li, Y.-L. Xia, S.-M. Ai, J. Liang, P. Sang, X.-L. Ji, S.-Q. Liu, *Int. J. Mol. Sci.* **2016**, *17*, 144.
- [66] M. L. Doyle, *Curr Opin Biotechnol* **1997**, *8*, 31-35.
- [67] J. E. Ladbury, B. Z. Chowdhry, *Chem. Biol.* **1996**, *3*, 791-801.

- [68] O. Cala, F. Guillièrre, I. Krimm, *Anal. Bioanal. Chem.* **2014**, *406*, 943-956.
- [69] J. Clayden, N. Greeves, S. Warren, P. Wothers, *Organic Chemistry*, **2001**.
- [70] S. M. Kelly, T. J. Jess, N. C. Price, *Biochim. Biophys. Acta - Proteins Proteomics* **2005**, *1751*, 119-139.
- [71] C. L. Meyerkord, *Molecular Biology 1278 Protein-Protein Interactions*, **n.d.**
- [72] S. Kelly, N. Price, *Curr. Protein Pept. Sci.* **2005**, *1*, 349-384.
- [73] M. Asmari, R. Ratih, H. A. Alhazmi, S. El Deeb, *Methods* **2018**, DOI 10.1016/j.ymeth.2018.02.003.
- [74] M. Mutter, S. Vuilleumier, *Angew. Chem. Int. Ed.* **1989**, *28*, 535-554.
- [75] M. Mutter, P. Dumy, P. Garrouste, C. Lehmann, M. Mathieu, C. Peggion, S. Peluso, A. Razaname, G. Tuchscherer, *Angew. Chemie Int. Ed. English* **1996**, *35*, 1482-1485.
- [76] S. Peluso, P. Dumy, C. Nkubana, Y. Yokokawa, M. Mutter, *J. Org. Chem.* **1999**, *64*, 7114-7120.
- [77] D. N. Posnett, H. Mcgrath, P. Tam, *Biochemistry* **1988**, *263*, 1719-1725.
- [78] J. P. Tam, *Proc. Natl. Acad. Sci.* **1988**, *85*, 5409-5413.
- [79] R. Lo-Man, S. Vichier-Guerre, S. Bay, E. Dériaud, D. Cantacuzène, C. Leclerc, *J. Immunol.* **2001**, *166*, 2849-2854.
- [80] R. Lo-Man, S. Vichier-Guerre, R. Perraut, E. Dériaud, V. Huteau, L. BenMohamed, O. M. Diop, P. O. Livingston, S. Bay, C. Leclerc, *Cancer Res.* **2004**, *64*, 4987-4994.
- [81] C. Geraci, G. M. L. Consoli, E. Galante, E. Bousquet, M. Pappalardo, A. Spadaro, *Bioconjug. Chem.* **2008**, *19*, 751-758.

- [82] P. S. Ghosh, A. D. Hamilton, *J. Am. Chem. Soc.* **2012**, *134*, 13208-13211.
- [83] J. M. Holub, *Drug Dev. Res.* **2017**, *78*, 268-282.
- [84] A. P. Silverman, A. M. Levin, J. L. Lahti, J. R. Cochran, **2009**, *385*, 1064-1075.
- [85] O. Longin, H. van de Langemheen, R. M. J. Liskamp, *Bioorganic Med. Chem.* **2017**, *25*, 5008-5015.
- [86] M. Hijnen, D. J. van Zoelen, C. Chamorro, P. van Gageldonk, F. R. Mooi, G. Berbers, R. M. J. Liskamp, *Vaccine* **2007**, *25*, 6807-6817.
- [87] P. R. Werkhoven, H. Van De Langemheen, S. Van Der Wal, J. a W. Kruijtzer, R. M. J. Liskamp, *J. Pept. Sci.* **2014**, *20*, 235-239.
- [88] P. R. Werkhoven, M. Elwakiel, T. J. Meuleman, H. C. Quarles van Ufford, J. A. W. Kruijtzer, R. M. J. Liskamp, *Org. Biomol. Chem.* **2015**, *14*, 701-10.
- [89] B. Fabre, J. Pícha, V. Vaněk, M. Buděšínský, J. Jiráček, *Molecules* **2015**, *20*, 19310-19329.
- [90] T. A. Hill, N. E. Shepherd, F. Diness, D. P. Fairlie, *Angew. Chemie - Int. Ed.* **2014**, *53*, 13020-13041.
- [91] A. Tapeinou, M. T. Matsoukas, C. Simal, T. Tselios, *Biopolymers* **2015**, *104*, 453-461.
- [92] A. Zorzi, K. Deyle, C. Heinis, *Curr. Opin. Chem. Biol.* **2017**, *38*, 24-29.
- [93] O. Al Musaimi, D. Al Shaer, B. de la Torre, F. Albericio, *Pharmaceuticals* **2018**, *11*, 42.
- [94] C. A. Lipinski, F. Lompardo, B. W. Dominy, P. J. Feeney, *Adv. drug Deliv. Rev.* **2001**, *46*, 3-26.
- [95] C. A. Lipinski, *Drug Discov. Today Technol.* **2004**, *1*, 337-341.

- [96] E. Ruoslahti, M. D. Pierschbacher, **1986**, *44*, 517-518.
- [97] C. Mas-Moruno, F. Rechenmacher, H. Kessler, *Anticancer. Agents Med. Chem.* **2010**, *10*, 753-768.
- [98] R. Stupp, M. Picard, M. Weller, *Lancet Oncol.* **2014**, *15*, 585-586.
- [99] T. Moroishi, T. Hayashi, W. W. Pan, Y. Fujita, M. V. Holt, J. Qin, D. A. Carson, K. L. Guan, *Cell* **2016**, *167*, 1525-1539.
- [100] Z. Zhang, Z. Lin, Z. Zhou, H. C. Shen, S. F. Yan, A. V Mayweg, Z. Xu, N. Qin, J. C. Wong, Z. Zhang, et al., **2014**.
- [101] M. D. Galsky, N. J. Vogelzang, P. Conkling, J. Polzer, S. Roberson, J. R. Stille, M. Saleh, D. Thornton, *Clin. Cancer Res.* **2014**, *20*, 3581-3588.
- [102] B. Wu, E. Y. T. Chien, C. D. Mol, G. Fenalti, W. Liu, R. Abagyan, A. Brooun, P. Wells, F. C. Bi, J. Damon, et al., **2010**, *330*, 1066-1071.
- [103] R. Salgia, J. R. Stille, R. W. Weaver, M. McCleod, O. Hamid, J. Polzer, S. Roberson, A. Flynt, D. R. Spigel, *Lung Cancer* **2017**, *105*, 7-13.
- [104] J. E. Crowe, *Cell Host Microbe* **2017**, *22*, 193-206.
- [105] R. J. Kuhn, W. Zhang, M. G. Rossmann, **2002**, 717-725.
- [106] G. Fibriansah, J. L. Tan, S. A. Smith, A. R. de Alwis, T. S. Ng, V. A. Kostyuchenko, K. D. Ibarra, J. Wang, E. Harris, A. de Silva, et al., *EMBO Mol. Med.* **2014**, *6*, 358-371.
- [107] L. Milroy, T. N. Grossmann, S. Hennig, L. Brunsveld, C. Ottmann, *Chem. Rev.* **2014**, *114*, 4695-4748.
- [108] O. Longin, M. Hezwani, H. van de Langemheen, R. M. J. Liskamp, *Org. Biomol. Chem.* **2018**, *16*, DOI 10.1039/C8OB01104D.
- [109] S. Fahs, Y. Patil-Sen, T. J. Snape, *ChemBioChem* **2015**, *16*, 1840-1853.

- [110] J. A. Robinson, S. C. Shankaramma, P. Jetter, U. Kienzl, R. A. Schwendener, J. W. Vrijbloed, D. Obrecht, *Bioorganic Med. Chem.* **2005**, *13*, 2055-2064.
- [111] A. Bunschoten, J. H. Ippel, J. A. W. Kruijtzer, L. Feitsma, C. J. C. De Haas, R. M. J. Liskamp, J. Kemmink, *Amino Acids* **2011**, *40*, 731-740.
- [112] M. Rubens, V. Ramamoorthy, A. Saxena, N. Shehadeh, S. Appunni, *J. Immunol. Res.* **2015**, *2015*, DOI 10.1155/2015/560347.
- [113] M. Sela, R. Arnon, B. Schechter, M. Sela, R. Arnon, **2002**, *7*, 664-673.
- [114] P. Werkhoven, Protein Mimics by Molecular Scaffolding of Peptides, **2016**.
- [115] G. R. Owen, J. A. Channell, V. T. Forsyth, M. Haertlein, E. P. Mitchell, A. Capovilla, M. Papathanasopoulos, N. M. Cerutti, *Biochemistry* **2016**, *55*, 2227-2237.
- [116] E. a Berger, T. R. Fuerst, B. Moss, *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 2357-61.
- [117] R. L. Garlick, R. J. Kirschner, F. M. Eckenrode, W. G. Tarpley, C. S. Tomich, *AIDS Res. Hum. Retroviruses* **1990**, *6*, 465-479.
- [118] S. E. Ryu, P. D. Kwong, a Truneh, T. G. Porter, J. Arthos, M. Rosenberg, X. P. Dai, N. H. Xuong, R. Axel, R. W. Sweet, *Nature* **1990**, *348*, 419-426.
- [119] E. Fenouillet, B. Clerget-Raslain, J. C. Gluckman, D. Guétard, L. Montagnier, E. Bahraoui, *J. Exp. Med.* **1989**, *169*, 807-822.
- [120] H. Yamaguchi, M. Miyazaki, *Biomolecules* **2014**, *4*, 235-251.
- [121] E. García-Fruitós, E. Vázquez, C. Díez-Gil, J. L. Corchero, J. Seras-Franzoso, I. Ratera, J. Veciana, A. Villaverde, *Trends Biotechnol.* **2012**, *30*, 65-70.
- [122] T. Zhang, X. Xu, L. Shen, *Acta Biochim. Biophys. Sin. (Shanghai).* **2009**,



41, 1044-1052.

- [123] R. R. Burgess, *Chapter 17 Refolding Solubilized Inclusion Body Proteins*, Elsevier Inc., **2009**.
- [124] P. Saha, B. Barua, S. Bhattacharyya, M. M. Balamurali, W. R. Schief, D. Baker, R. Varadarajan, *Biochemistry* **2011**, *50*, 7891-7900.
- [125] M. S. Osburne, E. a. Neidhardt, J. E. Godoy, M. R. Van Schravendijk, T. H. Grossman, *J. Immunol. Methods* **1999**, *224*, 19-24.
- [126] P. Graceffa, A. Jancsó, K. Mabuchi, *Arch. Biochem. Biophys.* **1992**, *297*, 46-51.
- [127] V. S. Alves, D. C. Pimenta, E. Sattlegger, B. A. Castilho, *Biochem. Biophys. Res. Commun.* **2004**, *314*, 229-234.
- [128] R. Varadarajan, D. Sharma, K. Chakraborty, M. Patel, M. Citron, P. Sinha, R. Yadav, U. Rashid, S. Kennedy, D. Eckert, et al., *J. Virol.* **2005**, *79*, 1713-1723.
- [129] E. Hochuli, W. Bannwarth, H. Döbeli, *Nat. Biotechnol.* **1988**, *6*, 1321-1325.
- [130] N. Cerutti, B. V. Mendelow, G. B. Napier, M. A. Papathanasopoulos, M. Killick, M. Khati, W. Stevens, A. Capovilla, *J. Biol. Chem.* **2010**, *285*, 25743-25752.
- [131] N. Sreerama, R. W. Woody, *Anal. Biochem.* **2000**, *287*, 252-260.
- [132] T. Rehm, R. Huber, T. A. Holak, *Structure* **2002**, *10*, 1613-1618.
- [133] G. S. Rule, T. K. Hitchens, *Fundamentals of Protein NMR Spectroscopy*, Springer, Dordrecht, The Netherlands, **2006**.
- [134] K. H. Khan, *Adv. Pharm. Bull.* **2013**, *3*, 257-263.
- [135] T. K. Kim, J. H. Eberwine, *Anal. Bioanal. Chem.* **2010**, *397*, 3173-3178.

- [136] I. Campbell, *Anaesth. Intensive Care Med.* **2007**, *8*, 163-167.
- [137] E. Selcuk Unal, R. Zhao, A. Qiu, I. D. Goldman, *Biochim. Biophys. Acta - Biomembr.* **2008**, *1778*, 1407-1414.
- [138] Y. Li, L. Luo, D. Y. Thomas, C. Y. Kang, *Virology* **1994**, *204*, 266-278.
- [139] W. Guo, B. Cleveland, T. M. Davenport, K. K. Lee, S. L. Hu, *Protein Expr. Purif.* **2013**, *90*, 34-39.
- [140] Z. Wang, C. Lorin, M. Koutsoukos, D. Franco, B. Bayat, Y. Zhang, A. Carfi, S. Barnett, F. Porter, *Vaccines* **2016**, *4*, 17.
- [141] C. I. Murphy, J. R. McIntire, D. R. Davis, H. Hodgdon, J. R. Seals, E. Young, *Protein Expr. Purif.* **1993**, *4*, 349-357.
- [142] G. E. Mulder, J. A. W. Kruijtzer, R. M. J. Liskamp, *Chem. Commun.* **2012**, *48*, 10007-10009.
- [143] G. Mulder, H. L. C. Quarles van Ufford, J. van Ameijde, A. J. Brouwer, J. a W. Kruijtzer, R. M. J. Liskamp, *Org. Biomol. Chem.* **2013**, *11*, 2676-84.
- [144] H. van de Langemheen, H. (Linda) C. Quarles van Ufford, J. A. W. Kruijtzer, R. M. J. Liskamp, *Org. Lett.* **2014**, *16*, 2138-2141.
- [145] C. Chamorro, J. a W. Kruijtzer, M. Farsaraki, J. Balzarini, R. M. J. Liskamp, *Chem. Commun. (Camb)*. **2009**, 821-823.
- [146] T. J. Meuleman, J. I. Dunlop, A. M. Owsianka, H. Van De Langemheen, A. H. Patel, R. M. J. Liskamp, *Bioconjug. Chem.* **2018**, *29*, 1091-1101.
- [147] A. Kovalová, R. Pohl, M. Vrabel, *Org. Biomol. Chem.* **2018**, DOI 10.1039/C8OB01617H.
- [148] A. A. H. Ahmad Fuaad, F. Azmi, M. Skwarczynski, I. Toth, *Molecules* **2013**, *18*, 13148-13174.

- [149] T. Zhou, L. Xu, B. Dey, A. J. Hessel, D. Van Ryk, S.-H. Xiang, X. Yang, M.-Y. Zhang, M. B. Zwick, J. Arthos, et al., *Nature* **2007**, *445*, 732-737.
- [150] G. M. Fischer, C. Jüngst, M. Isomäki-Krondahl, D. Gauss, H. M. Möller, E. Daltrozzo, A. Zumbusch, *Chem. Commun. (Camb)*. **2010**, *46*, 5289-5291.
- [151] J. McConathy, D. Zhou, S. E. Shockley, L. A. Jones, E. A. Griffin, H. Lee, S. J. Adams, R. H. Mach, *Mol. Imaging* **2010**, *9*, 329-342.
- [152] G. I. Tesser, I. C. Balvert-Geers, *Int. J. Pept. Protein Res.* **1975**, *7*, 295-305.
- [153] C. D. Rizzuto, R. Wyatt, N. Hernández-Ramos, Y. Sun, P. D. Kwong, W. A. Hendrickson, J. Sodroski, *Science (80-. )*. **1998**, *280*, 1949-1953.
- [154] G. E. Mulder, H. C. Quarles Van Ufford, J. Van Ameijde, A. J. Brouwer, J. A. W. Kruijtzer, R. M. J. Liskamp, *Org. Biomol. Chem.* **2013**, *11*, 2676-2684.
- [155] H. van de Langemheen, V. Korotkovs, J. Bijl, C. Wilson, S. S. Kale, C. Heinis, R. M. J. Liskamp, *ChemBioChem* **2017**, *18*, 387-395.
- [156] D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges, B. D. Sykes, *J. Biomol. NMR* **1995**, *5*, 67-81.
- [157] L. Whitmore, B. A. Wallace, *Nucleic Acids Res.* **2004**, *32*, 668-673.
- [158] D. G. Myszka, R. W. Sweet, P. Hensley, M. Brigham-Burke, P. D. Kwong, W. A. Hendrickson, R. Wyatt, J. Sodroski, M. L. Doyle, *Proc. Natl. Acad. Sci.* **2000**, *97*, 9026-9031.
- [159] S. Raghava, B. Barua, P. K. Singh, M. Das, L. Madan, S. Bhattacharyya, K. Bajaj, B. Gopal, R. Varadarajan, M. N. Gupta, *Protein Sci.* **2008**, *17*, 1987-1997.
- [160] S. G. Patching, *Biochim. Biophys. Acta - Biomembr.* **2014**, *1838*, 43-55.

[161] GE Healthcare Life Science, **2012**, 1-78.

[162] *GE Heal. Handbooks* **2009**, 49, 1-108.